Arnold Schwarzenegger
Governor

VALIDATION OF LATERAL BOUNDARY CONDITIONS FOR REGIONAL CLIMATE MODELS

PIER INTERIM PROJECT REPORT

Prepared For:
California Energy Commission
Public Interest Energy Research Program

Prepared By:
Tom Wigley (NCAR), Herbie Lee (UCSC)
Angela Pignotti (UCSC)

Insert image here…..

July 2009
CEC-500-XXXX-XXX

Prepared By:
Tom Wigley
NCAR, Boulder, CO

Herbie Lee
Angela Pignotti
UCSC, Santa Cruz, CA

Commission Contract No. Insert: #
Commission Work Authorization No: Insert: #

Prepared For:
Public Interest Energy Research (PIER)
California Energy Commission

Guido Franco
Contract Manager

Insert: Program Area Lead Name
Program Area Lead
Insert: Program Area Name

Insert: Office Manager Name
Office Manager
Insert: Office Name

Thom Kelly, Ph.D.
Deputy Director
ENERGY RESEARCH & DEVELOPMENT DIVISION

Melissa Jones
Executive Director

Acknowledgments

The authors thank Doug Nychka, NCAR, for initial guidance on some of the statistical aspects of this project. The AOGCM modeling groups are gratefully acknowledged for providing their climate simulation data through the Program for Climate Model Diagnosis and Intercomparison (PCMDI). We also acknowledge PCMDI for collecting and archiving these data, and the World Climate Research Programme's Working Group on Coupled Modelling for organizing the model data analysis activity. The CMIP3/AR4 multi-model data set is supported by the Office of Science, U.S. Department of Energy.

Please cite this report as follows:

Wigley, T.M.L., Pignotti, A.J. and Lee, H., 2009: *Validation of Lateral Boundary Conditions for Regional Climate Models.* California Energy Commission, PIER Energy-related Environmental Research Program. CEC-500-2009-xxx

Preface

The California Energy Commission's Public Interest Energy Research (PIER) Program supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The PIER Program conducts public interest research, development, and demonstration (RD&D) projects to benefit California.

The PIER Program strives to conduct the most promising public interest energy research by partnering with RD&D entities, including individuals, businesses, utilities, and public or private research institutions.

- PIER funding efforts are focused on the following RD&D program areas:
- Buildings End-Use Energy Efficiency
- Energy Innovations Small Grants
- Energy-Related Environmental Research
- Energy Systems Integration
- Environmentally Preferred Advanced Generation
- Industrial/Agricultural/Water End-Use Energy Efficiency
- Renewable Energy Technologies
- Transportation
-

Validation of Lateral Boundary Conditions for Regional Climate Models is the interim report for the [project name] project (contract number XXX-XX-XXX, work authorization number [insert #] or grant number [insert #]) conducted by the National Center for Atmospheric Research and the University of California at Santa Cruz. The information from this project contributes to PIER's Energy-Related Environmental Research Program.

For more information about the PIER Program, please visit the Energy Commission's website at www.energy.ca.gov/research/ or contact the Energy Commission at 916-654-4878.

Table of Contents

List of Figures

Figure 5.4: Seasonal movements of jet latitudes in the model (CCSM3.0) and reanalysis (NCEP). Given the difficulty in precisely defining jet location, the agreement is very good.

Figure 5.5: Seasonal changes in jet strengths in the model (CCSM3.0) and reanalysis (NCEP). The model seriously over-estimates the strength of the sub-tropical jet.

Figure 5.6: Vertical profiles of momentum fluxes across the western boundary at 130W. Results for both true fluxes (<uu>) and pseudo fluxes (<u><u>) are shown, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are m^2/s^2.)

Figure 5.7: Vertical profiles of heat fluxes across the western boundary at 130W. Results for both true fluxes (<uT>) and pseudo fluxes (<u><T>) are shown, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are K m/s.)

Figure 5.8: Vertical profiles of moisture fluxes across the western boundary at 130W. Results for both true fluxes (<uq>) and pseudo fluxes (<u><q>) are shown, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.)

Figure 5.9: Bias in the CCSM3.0 model simulation of DJF precipitation. The map shows the model-mean percentage error. Blue areas show where the model simulation is too wet, red areas are where the model simulation is too dry.

Figure 6.1: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing CCSM3.0 model results with observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

Figure 6.2: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (GFDL2.1) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

Figure 6.3: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (PCM) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

Figure 6.4: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (GISS-EH) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

Figure 6.5: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (MIROCmedres) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Three runs (using different initializations) are shown for the model – bottom three rows.

Figure 6.6: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

Figure 6.7: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (GFDL2.1) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

Figure 6.8: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (PCM) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

Figure 6.9: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (GISS-EH) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

Figure 6.10: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (MIROCmedres) and observed (ERA40 reanalysis) results. (Units are K m/s.). Three runs (using different initializations) are shown for the model.

Figure 6.11: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

Figure 6.12: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (GFDL2.1) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

Figure 6.13: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (PCM) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

Figure 6.14: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (GISS-EH) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

Figure 6.15: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (MIROCmedres) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

Figure 7.1: 20-year monthly mean zonal wind field for the ERA40 and NCEP R1 reanalyses, and the CCSM3.0 run a, c and e model simulations. Zonal wind fields are averaged for each January over 20 years, 1980-1999.

Figure 7.2: Mahalanobis Distance (MD) between the ERA40 and NCEP reanalyses, and CCSM3.0 20th century runs A, C and E corresponding to the fields in Figure 7.1. The MDs are computed with respect to ERA40. NCEP is indicated as a red circle; CCSM simulations are displayed in green.

Figure 7.3: 20-year monthly mean specific humidity fields for the ERA40 and NCEP R1

reanalyses, and the CCSM3.0 run a, c and e model simulations. Fields are averaged for each January over 20 years, 1980-1999.

Figure 7.4: Mahalanobis Distance (MD) between the ERA40 and NCEP reanalyses, and CCSM3.0 20th century runs A, C and E corresponding to the fields in Figure 7.4. The MDs are computed with respect to ERA40. NCEP is indicated as a red circle; CCSM simulations are displayed in green.

Figure 7.5: 20-year monthly mean temperature fields for the ERA40 and NCEP R1 reanalyses, and the CCSM3.0 run a, c and e model simulations. Fields are averaged for each January over 20 years, 1980-1999.

Figure 7.6: MDs corresponding to the fields in Figure 7.5.

Figure 7.7: True momentum fluxes.

Figure 7.8: See text for details.

Figure 7.9: MD results for data in Figure 7.7.

Figure 7.10: MD results for momentum pseudo flux data in Figure 7.8.

Figure 7.11a: July heat fluxes (cf. Figure 7.7).

Figure 7.11b: MD values for Figure 7.11a.

Figure 7.12a: July heat fluxes (cf. Figure 7.8).

Figure 7.12b: MDs for data in Figure 7.12a.

Figure 7.13a: July moisture pseudo fluxes.

Figure 7.13b: MDs for the data in Figure 7.13a.

Figure 7.14a: July moisture fluxes.

Figure 7.14b: MDs for the data in Figure 7.14a.

List of Tables

Table 1.1: AOGCMs considered in the study.

Table 2.1: Validation statistics used for ranking models. The variable used for ranking is annual precipitation for the globe and winter (DJF) precipitation for the western USA region (20–55N, 100–130W). The first numbers in each column are for the globe, while the second numbers are for western USA. The top three models for each statistic are shown in bold red type, while the worst three models in each case are shown in bold blue type. MODBAR refers to results for the model average. The score (column 1) is the number of top seven placings minus the number of bottom seven placings.

Table 3.1: Outlier analysis. For each model, the comparison statistic compares the model result for the normalized percentage change in annual precipitation with the average across all other models. For bias, this is model-i minus the average over the remaining 19 models. The analysis field is the whole globe.

Table 4.1: Comparison of zonal flow and fluxes at 130W with ENSO for reanalyses and CCSM3.0. Zonal flow and fluxes are characterized by PC1 of the monthly-mean time series over 1980 to 1999. ENSO is characterized by the Nino 3.4 SST time series. Numbers in the Table are correlation coefficients, followed in square brackets by the lag (positive lag means that the flow or flux lags behind ENSO) that gives the maximum correlation. The positive correlations for the flow/ENSO link indicate that zonal flows are stronger when Nino 3.4 SSTs are warmer. All flux correlations are shown as positive because the sign of the PC loading is arbitrary.

Table 4.2: ENSO performance metrics. Performance ranks only are given for RMSE and pattern correlation. The top 10 models are indicated by blue type. Cases where the model variability is too high (rather than too low) are shown as red in the last column.

Table 6.1: Observed and modeled momentum flux characteristics – maximum flux and latitude of the maximum. The MODEL MEAN results are to give a simple overview – it hides large inter-model differences. The first number is the latitude of the maximum (degrees N) and the second number is the maximum flux(m^2/s^2).

Table 6.2: Observed and modeled moisture flux characteristics – maximum fl;ux and latitude of the maximum. The first number is the latitude of the maximum (degrees N) and the second number is the maximum flux(kg m/ kg s).

Abstract

High-resolution Regional Climate Models (RCMs) are driven at their lateral boundaries by information from global models (usually coupled Atmosphere/Ocean General Circulation Models – AOGCMs). RCMs are generally driven by 6-hourly data from the AOGCM, and the spatial AOGCM data are interpolated to the boundaries of the RCM grid. When driven by observed (reanalysis) data, RCMs show high skill in their simulations of present-day climate within their domain, attributable largely to the improved resolution of surface boundary conditions (especially orography) relative to global models. For projections of future climate, however, when the RCM is driven by future climate-change output from an AOGCM, the skill of an RCM will depend to some degree on the skill of the AOGCM. For the best RCM results it is likely that these will be produced by the best driver AOGCMs. The question therefore arises as to how to decide what are the best AOGCMs.

There are different ways to assess the relative skill of different AOGCMs. We consider four methods here. First, we investigate how well different AOGCMs simulate present-day climate – better models are those that simulate present climate better. Second, we compare projections of future climate across a range of AOGCMs. We judge models whose projections differ greatest from the model-mean projections (outlier models) as least reliable. Third, we consider ENSO performance. Present and future climate over the California region is strongly linked to the El Nino/Southern Oscillation phenomenon (ENSO). Thus, AOGCMs that produce poorer simulations of ENSO should be judged less useful as RCM drivers.

As a fourth criterion we consider the western boundary fluxes directly. Climate and climate changes within an RCM domain must be dependent to a large degree on the fluxes of mass, momentum, heat and moisture into the domain along its western boundary. It is important, therefore, to assess how well AOGCMs can simulate present-day lateral boundary conditions. This provides a fourth criterion for selecting those AOGCMs that are best as RCM drivers. We consider both real fluxes calculated using 6-hourly data (viz. <uX> for variable 'X' where u is the westerly wind speed and <…> denotes a time average) and 'pseudo fluxes' defined by <u><X>, which require only monthly data for their calculation. We show that, in terms of their implications for validation of model fluxes, pseudo fluxes give the same results as real fluxes and so may be used as a replacement for real fluxes.

For the validation of boundary fluxes, we use the Mahalanobis Distance as a metric for determining how well a model matches the observations, and we develop statistical tests to determine whether model/observed differences are statistically significant.

We have assessed 20 models from the AR4/CMIP3 data base. For the validation, outlier and ENSO test criteria we are able to divide the models into three groups. Superior model are CCSM3.0, GFDL2.0, GFDL2.1, IPSL, MIROCmedres and HadCM3. Inferior models that cannot be recommended as RCM drivers are CNRM, FGOALS, GISS-EH, GISS-ER, INM and PCM.

Intermediate models are CCCMA, MRI, ECHO-G, BCCR, CSIRO, MIROChires, ECHAM5 and HadGEM1. We note that CCMA, MRI and ECHO-G are flux adjusted models, which may produce a favorable bias in their validation performance, so these should be used, if at all, with caution.

For direct validation of western boundary flux performance we have examined only CCSM3.0, GDFL2.1, PCM, GISS-EH and MIROCmedres. We find major errors in all of these models in their simulations of the strengths of the subtropical and polar jets – all models produce jets that are too strong. Moisture flux simulations are better. Here, MIROCmedres is the best model, followed in order by GFDD2.1, CCSM3.0, GISS-EH and PCM. The last two models here cannot be recommended as RCM drivers.

Keywords: AOGCMs, regional models, RCMs, boundary fluxes, model validation

1.0 Introduction

The primary tools for investigating the impacts of future climate change are Atmosphere/Ocean General Circulation Models (AOGCMs). In many climate impact areas, however, particularly agriculture and water resources, the analyses require higher resolution data than is generally available from these models. There are three standard methods for generating such high-resolution information, grouped under the name 'downscaling'. Statistical downscaling makes use of present-day relationships between AOGCM-scale data (typically around 300km by 300km) and higher resolution data (50km by 50km or less, and often at the single station level) and applies these to future simulations. Physical downscaling uses a higher-resolution (50km by 50km or less) limited-area model (a Regional Climate Model – RCM) effectively 'embedded' within the AOGCM to produce finer-scale detail. The RCM is driven by lateral boundary condition data from the AOGCM. Time-slice experiments use AOGCM sea-surface temperature (SST) data to drive a global atmospheric model (AGCM) that has much finer resolution than the original AOGCM. Hybrid schemes are also used, where the AOGCM drives an intermediate resolution AGCM, and lateral boundary data from the AGCM are used to drive the RCM. Further information and key literature references are given in Wigley (2004) and in the IPCC Fourth Assessment Report (AR4; Christensen and Hewitson, 2007).

In physical downscaling one has to make a choice as to which AOGCM or AOGCMs should be used to provide the (time-varying) lateral boundary conditions for driving an 'embedded' RCM. The primary criterion for choosing an AOGCM is its skill in simulating present-day climate. There is no standard way to assess model skill, and different skill criteria may give different results – i.e., a model that is skillful by one criterion may be less skillful by another.

We will consider four methods to assess skill and we will use a set of 20 AOGCMs from the AR4/CMIP3 data base archived at PCMDI, LLNL as candidate driver AOGCMs. The goal is to identify the most skillful models overall, and to identify those models that are particularly unskillful. Between the best and the worst there will be a number of models of intermediate skill.

The first skill criterion is the fidelity with which different AOGCMs simulate present-day climate. This involves choosing which climate variables to consider, which region to use for the skill assessment, and which comparison statistic to use to quantify skill, For the variable, we use the one that is arguably the most important for regional climate change assessments over the California/Western USA region (CA/WUSA), namely precipitation. As study regions we use (1) a latitude/longitude box, 20-55N, 100-130W that represents the western USA, and (2) the whole globe. For WUSA we use winter (DJF) precipitation. For the whole globe we consider annual rather than DJF precipitation because DJF is summer in the Southern Hemisphere. There are physical differences between the seasons that would make the hemispheric results for DJF less directly comparable.

The WUSA regional analysis represents, to some degree, the model's response to boundary fluxes around the box, and so is an indirect indicator of flux skill. Skill at the global level is a measure of the credibility of the model's internal physics. Using two regions helps to account for results in a single small region that might be skillful purely by chance.

The second skill criterion we use is a measure of the consistency of future climate projections for a particular model compared with other models. By analogy with, e.g., a group of marching soldiers in which one or a few are out of step with the rest, models that differ markedly from other models ('outliers') are less likely to be useful as driver AOGCMs for regional climate modeling. We carry out outlier analyses for the same two regions and variables as noted above.

The third criterion is ENSO. ENSO variability is one of the main factors in determining the mean state and variability of climate over the CA/WUSA region; and we will show that the inter-annual variability of fluxes into the CA/WUSA region is strongly correlated with ENSO, It is crucially important, therefore, that any chosen driver AOGCM has a credible ENSO component. We therefore examine a number of aspects of ENSO that together allow us to determine which models are best at simulating this key determinant of natural variability. We will consider ENSO spectra, sea surface temperature (SST) variability in the Nino 3.4 region, and the patterns of SST during ENSO warm events.

The fourth criterion is the fluxes themselves. We consider fluxes of mass, momentum, heat and moisture along a western boundary situated at 130W (230E). For future climate, the quality of any RCM simulation will depend both on the quality of the RCM, and on the quality and realism of the lateral boundary conditions. Clearly, if the fluxes into the RCM domain are incorrect in their magnitudes or spatial patterns, then the climate simulated within the domain is almost certain to be flawed. We compare model simulations of the western boundary fluxes with observed fluxes as defined by reanalyses (see below). That flux fidelity is an important consideration has been demonstrated by the Hadley Centre's (unpublished) work with their regional model 'PRECIS' (more details are given in Wigley, 2004).

1.1 Observed data

As noted above, we employ so-called 'reanalysis' data to provide the lateral boundary conditions, specified at 6-hourly intervals and interpolated appropriately in space and time. To date, most studies using reanalysis data have used information from the 'first-generation' NCEP-NCAR reanalysis (Kalnay et al., 1996). Recently, a new 'second-generation' reanalysis has become available, the ECMWF reanalysis 'ERA-40' (Simmons and Gibson, 2000). Reanalyses are not 'pure' observations, but synthesized atmospheric states, derived by reprocessing sequences of past weather observations from both *in situ* and satellite sources using model-based data assimilation techniques developed to initiate numerical weather forecasts (Trenberth and Olson, 1988; Bengtsson and Shukla, 1988). Reanalyses provide an internally-consistent view of past climate changes over the whole globe, uncontaminated by the changes in model physics that affect the results from operational analyses.

Reanalyses have well-recognized deficiencies, especially prior to the satellite era beginning in 1979 (see, e.g., Basist and Chelliah, 1997; Pawson and Fiorino, 1998; Santer et al., 1999; Trenberth et al., 2001). There is strong evidence, however, that the ERA-40 product is superior to the NCEP-NCAR product (Santer et al., 2004). For the purposes of the present proposal we will assume that the reanalyses represent our best estimate of the observed 'truth'. By using two different reanalyses, we will provide some insights into the validity of this assumption. Our analyses will only consider data from 1979; specifically, the interval 1980 through 1999 that is common to the observations and all models in the AR4/CMIP3 data base.

1.2 Choice of AOGCMs

We use results from a number of different AOGCMs as complied in the AR4/CMIP3 data base. We consider 20 of these models, as listed in Table 1.1 below. These are the AOGCMs in the MAGICC/SCENGEN data base (http://www.cgd.ucar.edu/cas/wigley/magicc/). An additional model (GISS AOM) is also listed. This is not in the SCENGEN data base, but it is one of the models for which ENSO analyses have been performed.

Table 1.1 has two indicators – red type and bold font – to distinguish certain model aspects. First, four of the models (indicated in red) are 'flux adjusted' (see Randall and Wood, 2007). In other words, ocean data have been adjusted to improve the consistency between atmosphere-to-ocean and ocean-to-atmosphere fluxes. This is important because such models often fit observations better, and this may be partly an artifact of the flux adjustment process. One should be wary of model validation results and apparent skill in these cases. Second, our ENSO assessments rely on the comprehensive analysis of AR4 models by AchutaRao and Sperber (2006). The AOGCMs included in this analysis are indicated by **bold**, larger-font type in Table 1.1. Only BCCR and ECHO-G are not considered in the ENSO analysis.

For the boundary flux work, our most complete analyses employ data from NCAR's CCSM3.0, which runs at T85 spatial resolution (about 1¼ deg. by 1¼ deg.), extracted from an ensemble of simulations driven by observed natural and anthropogenic forcings. These runs have been designed to simulate recent climate variations and are appropriate for comparison with post-1979 reanalysis data.

Table 1.1: AOGCMs considered in the study.

CMIP3 designator	Country	SCENGEN name	Name used in text
BCCR-BCM2.0	Norway	BCCRBCM2	BCCR
CGCM3.1(T47)	Canada	CCCMA-31	CCCMA-
CCSM3.0	USA	CCSM—30	CCSM3.0
CNRM-CM3	France	CNRM-CM3	CNRM
CSIRO-Mk3.0	Australia	CSIRO-30	CSIRO
ECHO-G	Germany/Korea	ECHO---G	ECHO-G
FGOALS-g1.0	China	FGOALS1G	FGOALS
GFDL-CM2.0	USA	GFDLCM20	GFDL2.0
GFDL-CM2.1	USA	GFDLCM21	GFDL2.1
GISS-EH	USA	GISS—EH	GISS-EH
GISS-ER	USA	GISS—ER	GISS-ER
INM-CM3.0	Russia	INMCM-30	INM
IPSL-CM4	France	IPSL_CM4	IPSL
MIROC3.2(hires)	Japan	MIROC-HI	MIROChires
MIROC3.2(medres)	Japan	MIROCMED	MIROCmedres
MRI-CGCM2.3.2	Japan	MRI-232A	MRI
ECHAM5/MPI-OM	Germany	MPIECH-5	ECHAM5
PCM1	USA	NCARPCM1	PCM
UKMO-HadCM3	UK	UKHADCM3	HadCM3
UKMO-HadGEM1	UK	UKHADGEM	HADGEM1
GISS-AOM	USA		

1.3 Fluxes versus Pseudo Fluxes

One of the primary reasons for the focus on CCSM3.0 is that this is the only model for which we could obtain 6-hourly data. For a correct calculation of lateral boundary fluxes, 6-hourly data must be used. For all other models in the AR4/CMIP3 data base, only monthly-mean data have been archived. True fluxes cannot be calculated with monthly data, so we must distinguish between true fluxes and what we refer to here as ‘pseudo fluxes’. If <…> is used to indicate a time mean using 6-hourly data, then the true flux for a quantity ‘X’, is …

$$F = \langle uX \rangle$$

Here, u is the westerly wind component, and X = ρ (density) for mass flux, X = u for momentum flux, X = T for heat flux, and X = q (specific humidity) for moisture flux. A pseudo flux can be defined as …

$$P = \langle u \rangle \langle X \rangle$$

True and pseudo fluxes differ because time variations in u and X are correlated. Pseudo fluxes, have the advantage that they can be calculated with monthly data – i.e., with pseudo fluxes we do not require 6-hourly data. One of the tasks we tackle with CCSM3.0 is to compare the true and pseudo fluxes. If the implications for model skill from pseudo fluxes are similar to those for real fluxes, then we can use the monthly data in the AR4/CMIP3 archive to assess the credibility of AOGCM fluxes for a range of AOGCMs. In any event, the use of pseudo fluxes is a valid model/observed comparison tool. The key here is to use the same variables and comparison metrics from the climate model (AOGCM) and the observations (reanalysis).

1.4 Choice of Spatial Domain

There are a number of papers that have carried out regional modeling studies over the western United States, or the whole (contiguous) United States (see Wigley, 2004). Most of these studies have used domains that do not follow latitude and longitude lines. For the current analyses, it is both physically desirable and analytically simpler to employ a latitude-longitude domain. In the PIRCS regional model inter-comparison (Takle et al., 1999), where the basic domain was the contiguous United States, one of the eight models compared (viz. HIRHAM) used a latitude-longitude domain (one that is consistent with the domains used by the other seven models in PIRCS, and consistent with domains used in studies concentrating on the western United States; see Takle et al., 1999). We have therefore chosen to use the HIRHAM PIRCS domain as the basic domain for the present work. This runs from 55N, 130W to 20N, 70W – see the rectangular box in Figure 1.1. Figure 1.1 also shows a larger domain proposed for the NARCCAP project (Mearns et al., 2004; see Appendix 1). The western boundary of the rectangular box in Figure 1.1 (at 130W) is that used in the present study for calculating fluxes.

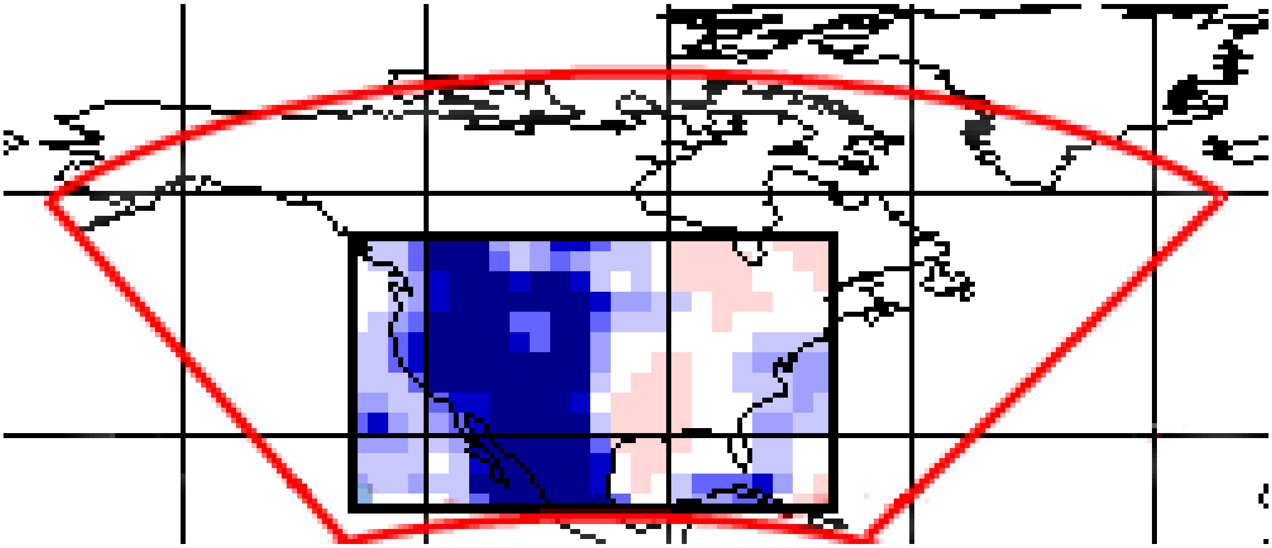

Figure 1.1: The rectangular latitude/longitude HIRHAM RCM region use to define the western flux boundary for the present study (130W), compared with the region proposed for the NARCCAP project (Mearns et al., 2004), red box.

2.0 Validation Analysis

For assessing model skill, our first method is to compare model simulations of present-day climate with observations. We do this considering two regions, WUSA and the globe. We use winter (DJF) precipitation as the key variable for WUSA and annual precipitation as the key variable for the global analysis. There are considerable uncertainties in quantifying both absolute skill and the relative skill of different models (see, e.g., Gleckler et al., 2008), so these results must be seen as just one of a number of possible skill assessments. As a validation variable we use precipitation in part because of its significance to California, and partly because precipitation is more difficult to model than, e.g., temperature. Models do less well in simulating precipitation than temperature, so using precipitation is a more stringent test of model skill.

The comparison statistics (i.e., skill metrics) used are: pattern correlation (r), root-mean-square error (RMSE), bias (B), and a bias-corrected RMSE (RMSE-corr). All statistics used here are employ cosine weighting to account for the changing area of grid boxes with latitude.

Bias is simply the difference, model minus observed, averaged over the chosen validation region. Of these four statistics, bias is probably the least important, since it is generally thought that biased models can still produce good information regarding future change, provided the bias is not too large. Bias may reflect incorrect baseline forcing (i.e., atmospheric composition and/or loadings of radiatively important species) in the runs used for validation, rather than a problem with model physics. Bias, however, can affect RMSE, which is why RMSE-corr results are given as a test statistic. RMSE-corr is the root-mean-square error after a correction is applied to the model-mean field to remove any overall bias. It is related to RMSE by

$$(\text{RMSE-corr})^2 = (\text{RMSE})^2 - B^2$$

Table 2.1 shows these statistics for all models in the SCENGEN data base. To rank models we have used a semi-quantitative skill score that rewards relatively good models and penalizes relatively bad models. Each model gets a score of +1 if it is in the top seven (top third approximately) for any statistic over the globe or over the WUSA, and a score of –1 if it is in the bottom seven. The maximum skill score is therefore +8, which would mean that the model was in the top seven for all four statistics over both regions. The worst possible score is –8. In Table 2.1, models are listed in order of their skill scores. Other skill scores could be devised – but the results for others that we have considered are similar.

Table 2.1 shows the clear superiority of the first three models – but note also that these three models are all flux adjusted. This may give them an advantage in a model validation exercise, so one should treat these results with caution. On the other hand, flux adjustment is not thought to be an issue for future climate change projections (see, e.g., Gregory and Mitchell, 1997), i.e., projections for a given model do not depend significantly on whether the model is flux adjusted or not. Nevertheless, if a flux-adjusted model validates well against present climate, as here, this may not be a good indicator of model quality. In these cases, other indicators of model quality should also be considered.

Not counting these three models, the best models are the two GFDL models, HadCM3, MIROCmedres and the CSIRO model. A few models are abysmally bad – GISS-ER, GISS-EH, CNRM and INM (the latter even though it is flux adjusted).

Table 2.1: Validation statistics used for ranking models. The variable used for ranking is annual precipitation for the globe and winter (DJF) precipitation for the western USA region (20–55N, 100–130W). The first numbers in each column are for the globe, while the second numbers are for western USA. The top three models for each statistic are shown in bold red type, while the worst three models in each case are shown in bold blue type. MODBAR refers to results for the model average. The score (column 1) is the number of top seven placings minus the number of bottom seven placings.

RANK (score)	FLUX ADJ?	MODEL	Pattern correlation	RMSE (mm/day)	Bias (mm/day)	RMSE-corr (mm/day)
1 (+8)	Yes	MRI	0.886/0.918	0.967/0.683	-0.084/+0.381	0.963/0.567
1 (+8)	Yes	CCCMA	0.888/0.927	0.949/0.798	-0.010/+0.592	0.949/0.536
3 (+7)	Yes	ECHO-G	0.910/0.894	0.864/1.098	+0.128/+0.793	0.854/0.760
4 (+6)		MIROCmedres	0.833/0.880	1.162/1.119	+0.035/+0.534	1.162/0.983
4 (+6)		GFDL2.0	0.868/0.857	1.099/1.136	+0.091/+0.803	1.095/0.804
6 (+4)		GFDL2.1	0.857/0.841	1.149/1.260	+0.215/+0.840	1.128/0.938
7 (+3)		HadCM3	0.858/0.850	1.256/1.155	+0.230/+0.647	1.235/0.957
7 (+3)		CSIRO	0.814/0.882	1.209/1.234	-0.161/+0.780	1.198/0.955
9 (-1)		IPSL	0.808/0.870	1.269/1.531	-0.090/+1.175	1.266/0.981
9 (-1)		ECHAM5	0.808/0.890	1.351/1.318	+0.247/+1.030	1.328/0.822
9 (-1)		MIROChires	0.800/0.894	1.340/1.532	+0.281/+0.892	1.311/1.245
12 (-2)		BCCR	0.793/0.804	1.311/1.243	+0.307/+0.629	1.275/1.072
13 (-3)		CCSM3.0	0.797/0.798	1.327/1.354	+0.160/+0.956	1.317/0.958
13 (-3)		FGOALS	0.816/0.757	1.226/2.121	+0.307/+1.729	1.187/0.969
13 (-3)		HadGEM1	0.797/0.929	1.614/1.349	+0.385/+0.437	1.568/1.276
13 (-3)		PCM	0.665/0.797	1.715/1.212	+0.343/+0.843	1.680/0.870
17 (-6)	Yes	INM	0.700/0.528	1.606/1.757	+0.116/+1.106	1.590/1.365
18 (-7)		GISS-EH	0.733/0.718	1.512/1.403	+0.340/+0.840	1.473/1.124
19 (-8)		CNRM	0.772/0.794	1.438/1.488	+0.540/+1.020	1.333/1.084
19 (-8)		GISS-ER	0.774/0.770	1.430/1.740	+0.297/+1.271	1.399/1.189
		MODBAR	0.910/0.919	0.870/1.090	+0.184/+0.865	0.850/0.663

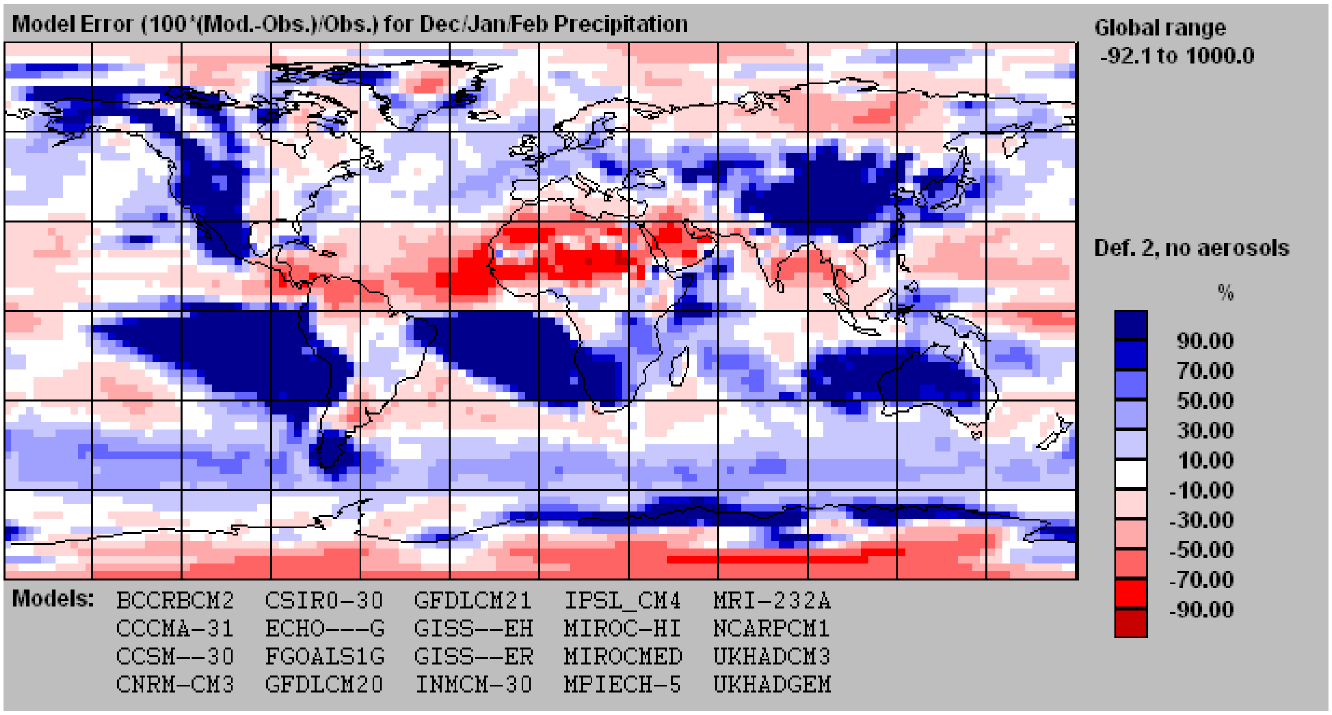

Figure 2.1: Bias in model simulations of DJF precipitation. The map shows the model-mean percentage error for the listed models. Blue areas show where model simulations are too wet, red areas are where model simulations are too dry. Note the serious model-mean positive bias over the western USA. This is common to all models (see Table 2.1)

An interesting result (now well known) is that averaging all the model results together (MODBAR in Table 2.1) leads to better overall validation performance than any individual model. Of course, this is not relevant for the present purpose. More relevant is the fact that all models have a positive bias over WUSA – they rain too much. The DJF bias ranges from 0.38 to 1.73 mm/day (about 5.5 to 24.8 inches/year) – see detail in Figures 2 and 3. The reasons for this are still uncertain, but the pattern of the bias suggests it is something to do with orography. As we will show below, these precipitation errors do not seem to be related to any moisture flux bias

Note also that models that perform well in terms of global statistics generally perform well over the much smaller western USA region. Models with high regional bias, however, may still perform adequately with the other statistics – CCSM3.0 and the two GFDL models are examples.

The biases shown in Figure 2.1 compare AOGCM control runs with the CMAP precipitation data base, which has the same resolution as the SCENGEN AOGCM data base (viz. 2.5 by 2.5 deg. latitude/longitude). The results here are consistent with other analyses of precipitation bias. An example (from Caldwell et al., 2009) is shown in Figure 2.2, which compares CCSM3 data with two much higher resolution precipitation data bases. (The extra spatial detail in Figure 2.2 comes from the high resolution observations, and would appear similarly if these data were compared with the lower-resolution CMAP data.) Caldwell et al. note that similar positive precipitation biases are common to many regional models, implying that the AOGCM errors are not simply due to orography, but must reflect flaws in model physics as well.

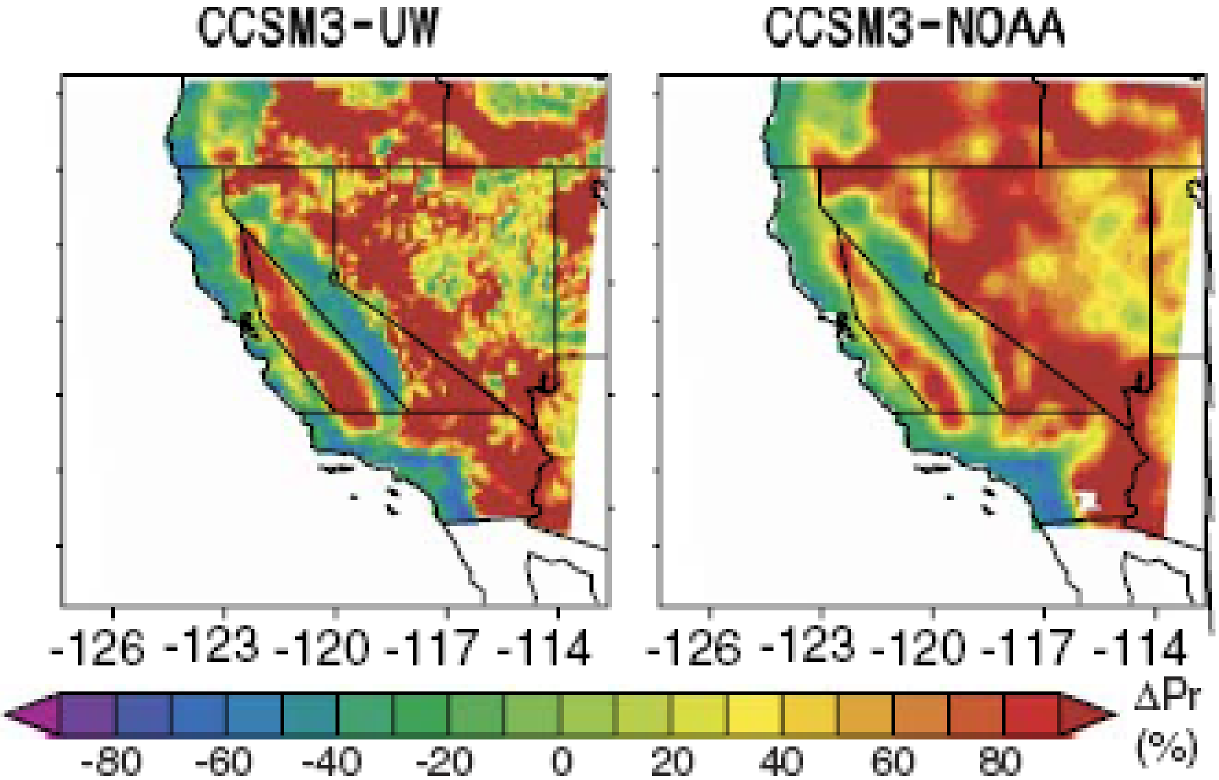

Figure 2.2: Comparison of model (CCSM3) and observed annual precipitation for two different high-resolution precipitation data bases (from Caldwell et al., 2009). The maps show results for (100*(Model minus Observed)/Observed).

3.0 Outlier Analysis

In selecting models it is also useful to compare model simulations of future climate. This is a way of factoring in the convergence criterion proposed by Giorgi and Mearns (2002) – these authors suggest that models that are less similar to other models (outliers) should be down-weighted. The analysis here uses all 20 models in the SCENGEN data base. We consider all 20 models even though it has already been noted in the previous section that some models (GISS-ER, GISS-EH, CNRM, INM) perform so poorly in validation that they almost certainly should not be used as RCM drivers.

To determine possible outliers we use normalized annual precipitation projections over the globe. The projections are in percentage terms, so it is percentage changes that we are comparing here. For each model we compare its projection with the average projection for all other models. In other words, if 'n' models are being considered, the normalized percentage changes for model 'i' are compared with the average of changes over all n-1 remaining models.

Note that, for 'all other models', this is the average of the percentage changes, which is not the same as (although generally similar to) the average absolute change expressed as a percentage. We use normalized projections (i.e., percentage change per degree C global-mean warming) to remove inter-model differences arising from differences in the climate sensitivity. We use the same comparison statistics that were used in the validation exercise – pattern correlation, RMSE, bias, and bias-corrected RMSE. Results are given in Table 3.1.

Based on this outlier analysis, there are a number of models that are strikingly different from other models in terms of their response patterns (based on the pattern correlation and bias-corrected RMSE results), viz., CNRM, GISS-ER, PCM, CSIRO, MRI and ECHAM5. Two of these (GISS-ER and CNRM) have already been identified as poor models on the basis of the validation analysis. It is interesting that two of the worst models based on this outlier analysis (ECHO-G and MRI) are among the best in terms of validation skill. Other models that show conflicting validation/outlier results are CSIRO, FGOALS and HadGEM1.

Table 3.1: Outlier analysis. For each model, the comparison statistic compares the model result for the normalized percentage change in annual precipitation with the average across all other models. For bias, this is model-i minus the average over the remaining 19 models. The analysis field is the whole globe.

MODEL	Pattern correl. [rank]	RMSE (%)	Bias (%)	RMSE-corr (%)
BCCR	0.480 [6]	6.873	0.515	6.854
CCCMA	0.608 [1]	5.737	-0.074	5.736
CCSM3.0	0.319 [15]	8.810	1.093	8.754
CNRM	0.260 [18]	8.243	0.271	8.239
CSIRO	0.291 [17]	9.548	0.709	9.521
ECHO-G	0.293 [16]	8.709	-0.759	8.676
FGOALS	0.513 [4]	8.647	-1.145	8.571
GFDL2.0	0.424 [7]	10.307	0.604	10.289
GFDL2.1	0.414 [9]	11.107	-2.058	10.914
GISS-EH	0.394 [12]	7.895	0.609	7.871
GISS-ER	0.124 [19]	24.100	0.245	24.099
INM	0.408 [10]	7.186	0.274	7.161
IPSL	0.422 [8]	10.000	-0.983	9.952
MIROChires	0.497 [5]	5.665	0.632	5.630
MIROCmedres	0.588 [2]	5.700	-0.121	5.699
MRI	0.350 [14]	15.456	0.960	15.426
ECHAM5	0.369 [13]	10.679	0.353	10.673
PCM	-0.099 [20]	15.363	0.914	15.336
HadCM3	0.404 [11]	10.149	-0.838	10.114
HadGEM1	0.525 [3]	6.513	-0.021	6.513

4.0 ENSO Assessment

ENSO is a quasi-periodic variation in sea surface temperatures in the tropical Pacific, which affects and is affected by the longitudinal Walker Circulation. ENSO variations are characterized by a normalized form of the pressure difference between Darwin and Tahiti (the Southern Oscillation Index) or by SST variations in the Nino 3 region or the Nino 3.4 region (see Figure 4.1). We use the SST indices here.

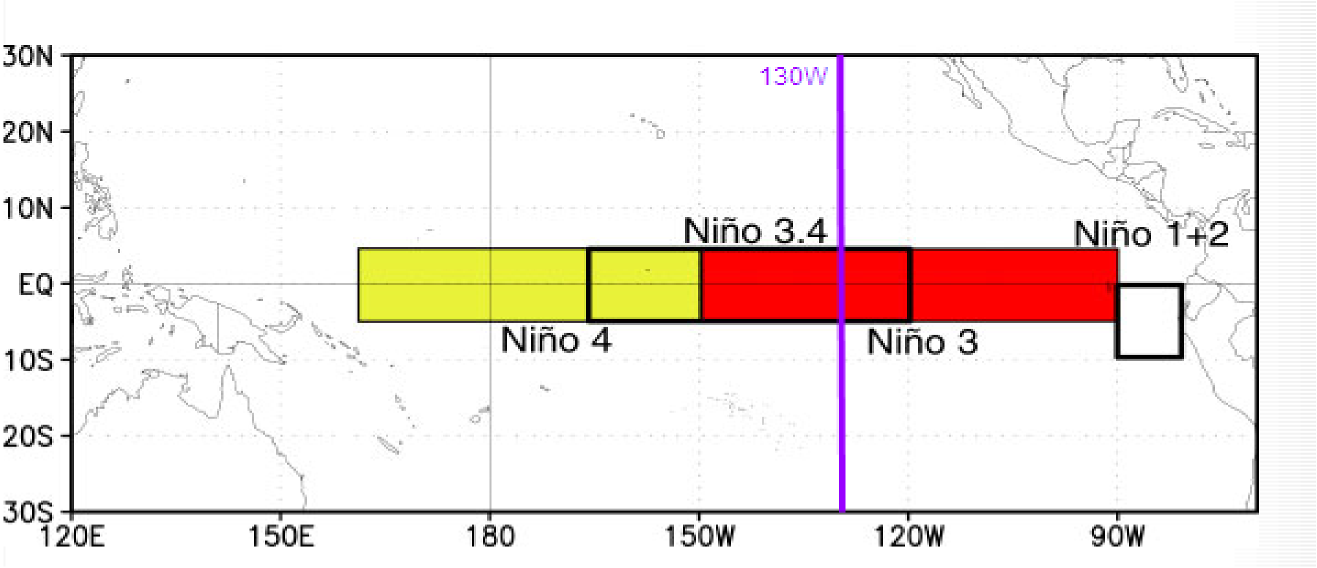

Figure 4.1: Sea surface temperature regions used for characterizing ENSO variations. The longitude (130W = 230E) used for determining easterly fluxes is shown by the vertical line.

ENSO is known to have a statistically significant effect on winter precipitation in California. During El Ninos (warm SSTs in the tropical Pacific), precipitation is enhanced, especially in southern California. In La Ninas (cool events), the signal is in the opposite direction and is less strong, but still detectable in southern California. A key criterion in choosing models for driving RCMs therefore should be the model's skill in simulating ENSO. More to the point, the key factor should be a model's skill in projecting future changes in ENSO. As we have no way to judge such skill, we rely on the assumption that skill in simulating present-day ENSOs is a guide to and prerequisite for skill in projections of change.

ENSO variations also appear to be an important determinant of changes in the western boundary fluxes, so this is another (and possibly more important) reason why RCM driver models should be those that are skillful in simulating ENSO.

4.1 ENSO/flux relationships

ENSO/flux relationships for both observations and CCSM3.0 are shown below in Figures 4.2 to 4.6. We characterize the fluxes by using the first principal component (PC1) of the monthly time series of the latitude-height patterns at 130W (over 1980 to 1999). Figure 4.2 shows ERA40 and CCSM3.0 results for the zonal flow.

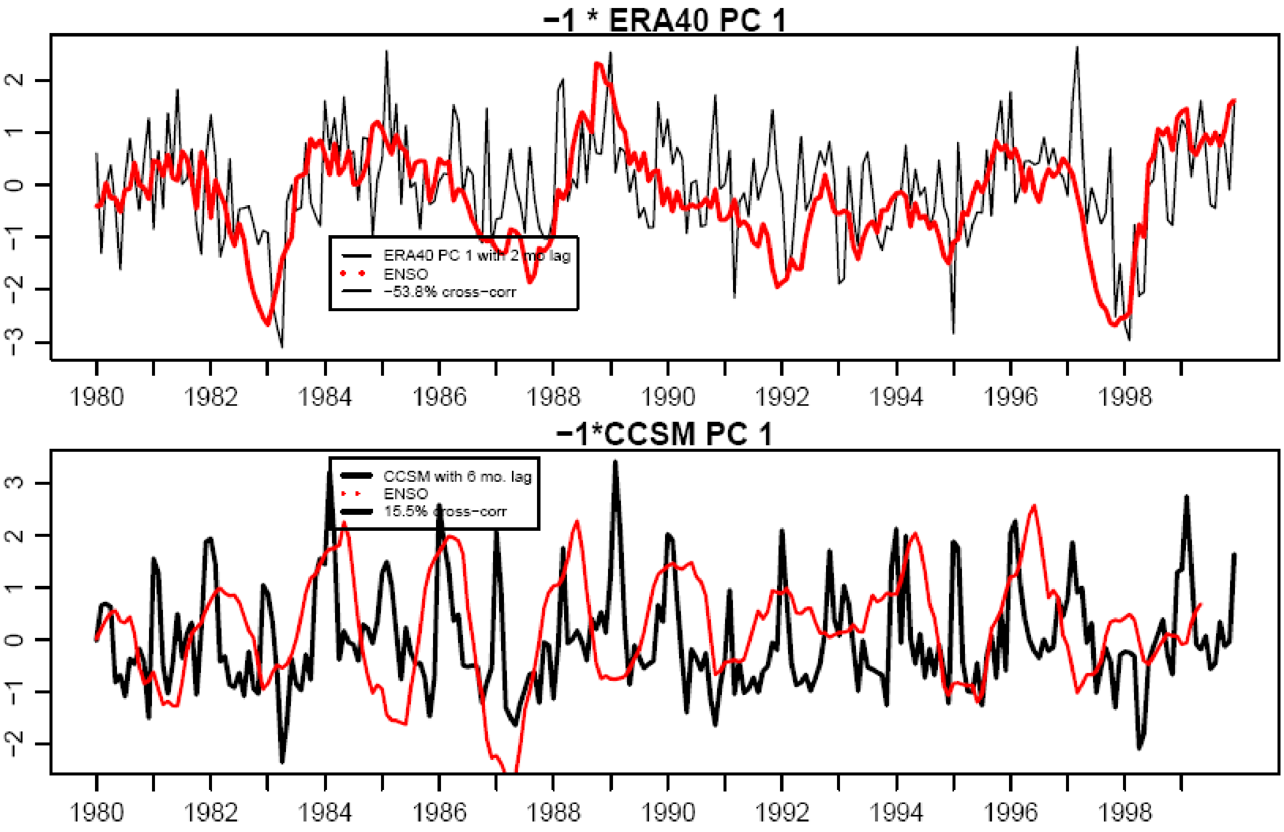

Figure 4.2: Comparison of zonal flow PC1 time series with ENSO (Nino 3.4 SSTs) time series for ERA40 and CCSM3.0. PC1 time series have been multiplied by –1 to facilitate comparison. For ERA40, the flow PC1 is lagged by 2 months to maximize the correlation. For CCSM3.0 the corresponding maximizing lag is 6 months.

There is a clear relationship between zonal flow and ENSO in the observations, maximized when the flow PC1 lags 2 months behind the ENSO series. The correlation between the PC1 and ENSO time series is $r = -0.538$, but the sign here is arbitrary. When flow magnitude is looked at directly, stronger flows are associated with warmer SSTs.

Can AOGCMs emulate this relationship? The answer appears to be “not well”. The bottom panel in Figure 4.2 shows the same relationship using CCSM3.0 data. In CCSM3.0, the ENSO time series necessarily differs from the observed ENSO time series, not only because the model world is essentially in a different universe, but also because CCSM3.0 does not simulate the ENSO power spectrum very well. The Figure shows that CCSM3.0’s ENSO has a very strong (and unrealistic) 2-year cycle (of which more will be said below). Furthermore, in the model world there is a much weaker link between ENSO and the zonal flow ($r = 0.155$), maximized at an unrealistic lag of 6 months.

Results for momentum flux <uu> are shown in Figure 4.3 using data from both the ERA40 and NCEP reanalyses, compared with CCSM3.0. (Note that CCSM3.0 is the only model for which we have 6-hourly data, and so is the only model where we can do a full flux comparison.) The results for the two reanalyses are virtually identical. Here, the maximum observed correlation occurs (as for the zonal flow) when the flux PC1 lags 2 months behind the SSTs ($r = 0.623$ for NCEP, $r = 0.632$ for ERA40). For

CCSM3.0, the relationship is much weaker (r = 0.286), maximized at zero lag. In terms of explained variance (r^2), the observed value of 40% (ERA40) drops to only 8% in the model.

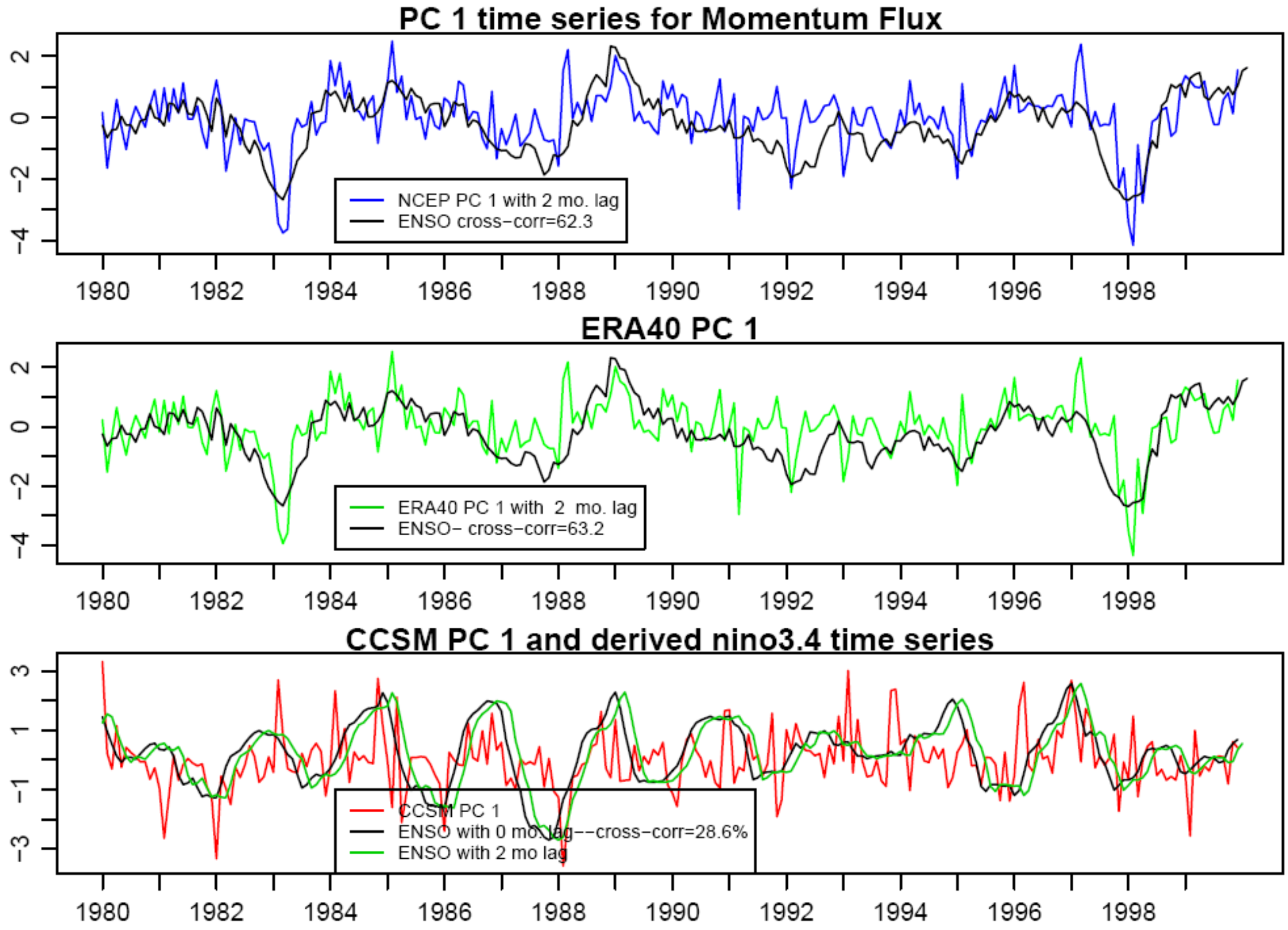

Figure 4.3: Comparison of momentum flux and ENSO variations for two reanalyses (NCEP and ERA40) and CCSM3.0.

Lag relationships between the momentum flux PC1 time series and Nino 3.4 SSTs are shown in Figure 4.4.

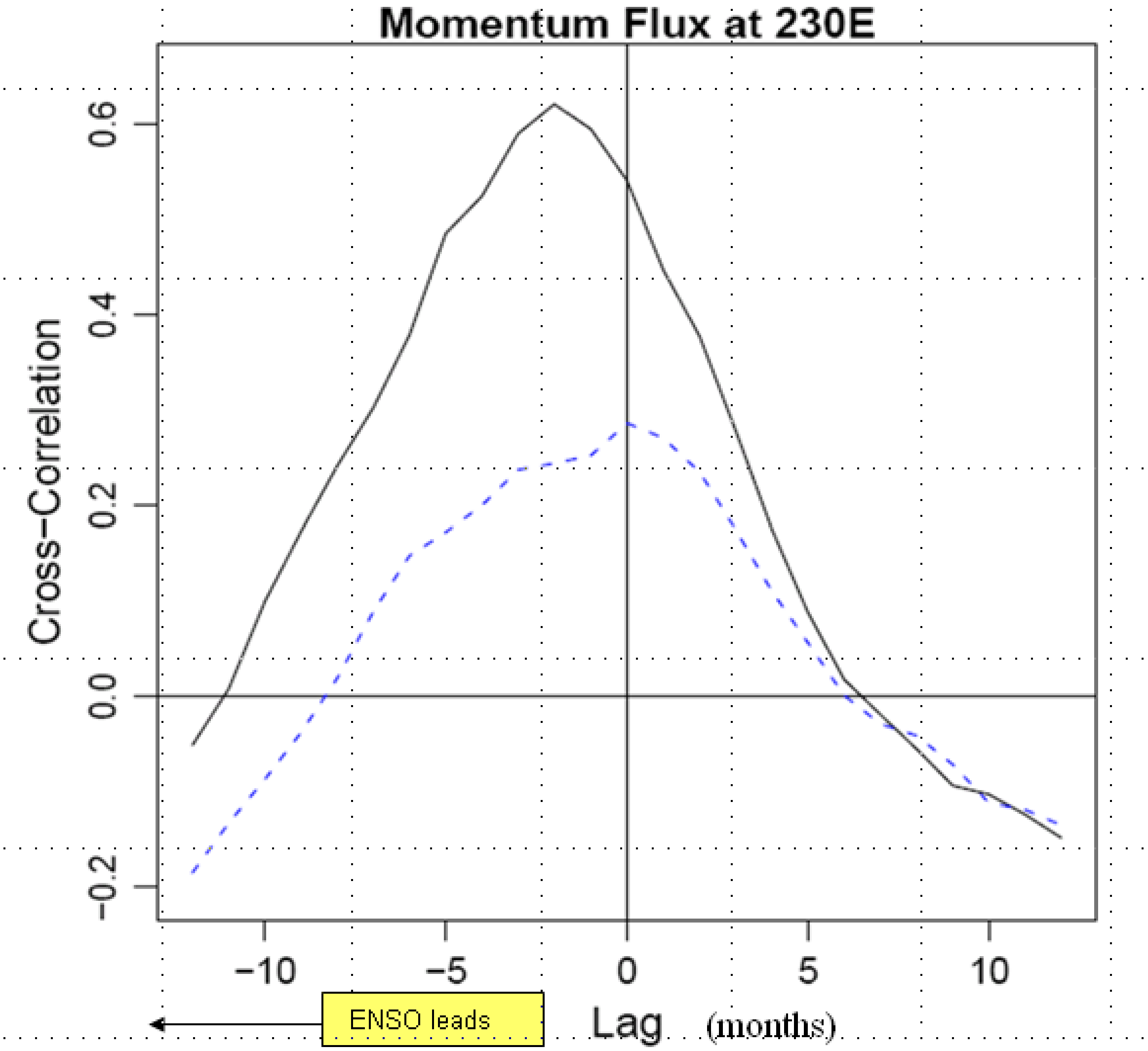

Figure 4.4: Lag correlations between momentum flux (PC1) and ENSO (Nino 3.4 SSTs) variations. Results are for ERA40 observed data (black line) and CCSM3.0 data (blue dashed line).

The reason for the ENSO/zonal wind and ENSO/momentum flux correlations almost certainly arises from the thermal wind relationship. Inter-annual SST variability is larger at low latitudes, so it is likely that the south-to-north temperature gradient will vary in parallel with tropical SSTs – i.e., in parallel with ENSO. The stronger this gradient, the larger will be the thermal wind, and, hence, the stronger will be the jet and the magnitude of the PC1 loading for zonal wind and momentum flux. The issue is complicated by the fact that there are two jets (see below). Further testing is required for this hypothesis.

Results for heat fluxes are shown in Figure 4.5. As heat flux involves absolute temperature, the spatial variations in temperature are much less than the variations in wind speed, so it is the spatial speed variations that will dominate the heat flux pattern. The results for heat flux are therefore similar to those for momentum flux: i.e., the flux/ENSO link is much weaker in the model than in the observations, and is maximized at zero lag compared with 2 months for the observations. There is essentially no difference between the two reanalyses.

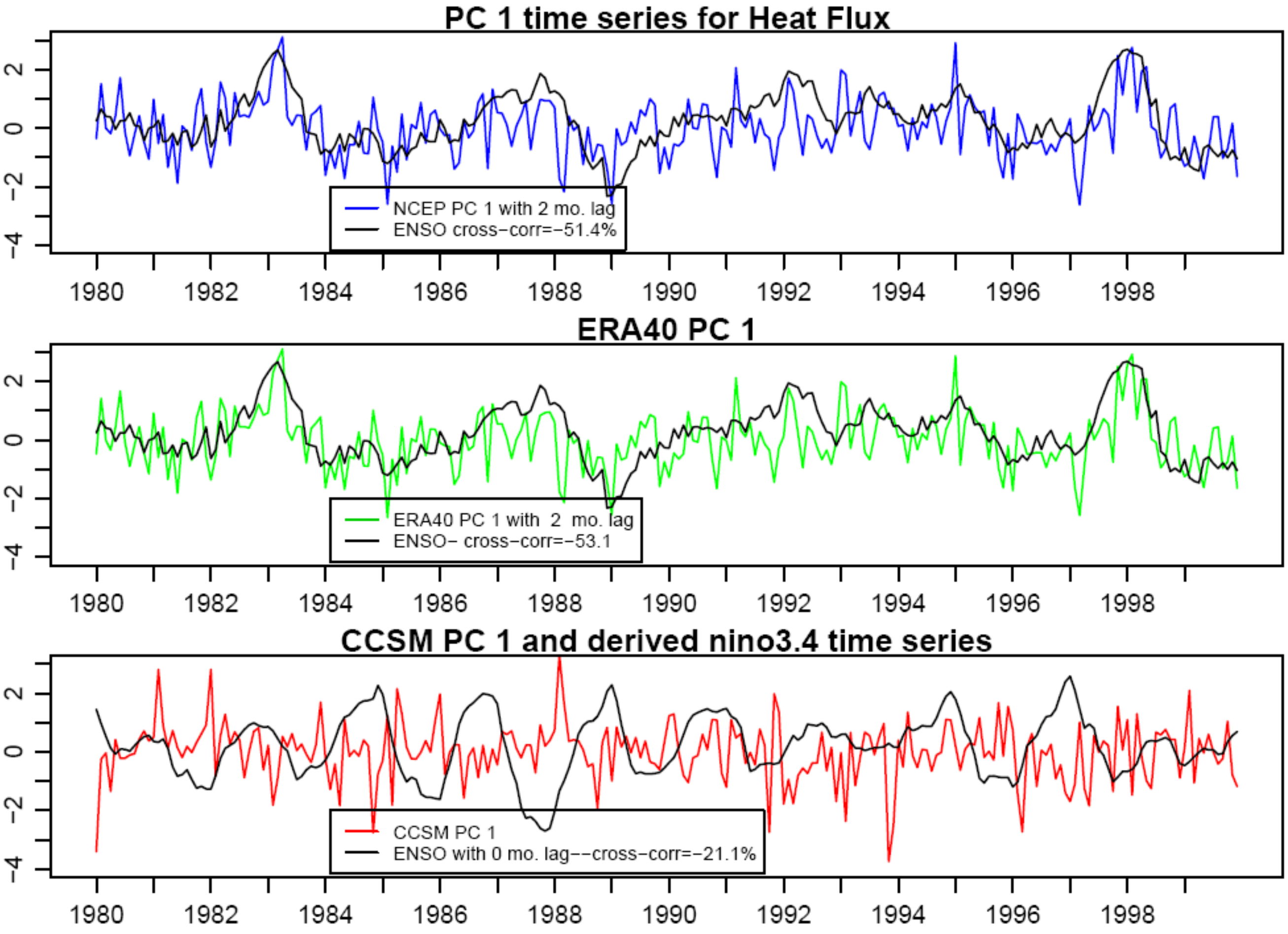

Figure 4.5: Comparison of heat flux and ENSO variations for two reanalyses (NCEP and ERA40) and CCSM3.0.

We can speculate what might hold for moisture fluxes. Here, atmospheric moisture content is much greater at low elevations, and is almost zero at the height of the jet. Moisture fluxes are therefore virtually independent on the strength of the jet, and, since it is jet strength that determines the ENSO links to zonal flow, momentum flux and heat flux, we might expect a much weaker link for moisture flux. On the other hand, lower atmosphere moisture content is strongly dependent on tropical SSTs, so there must be a direct link between moisture content (and, hence, moisture flux) and ENSO, independent of the jet. ENSO is therefore still likely to be an important factor in determining model skill for moisture flux. ENSO/moisture flux comparisons are shown in Figure 4.6, and results for all relationships are summarized in Table 4.1.

As with the zonal flow and fluxes, there is a statistically significant relationship between moisture flux and SSTs in the observations, but it is considerably weaker than for the other variables. Correlations are maximized at 2 months lag (which is somewhat surprising). In the model, the flux/ENSO link is much weaker, and maximized at zero lag.

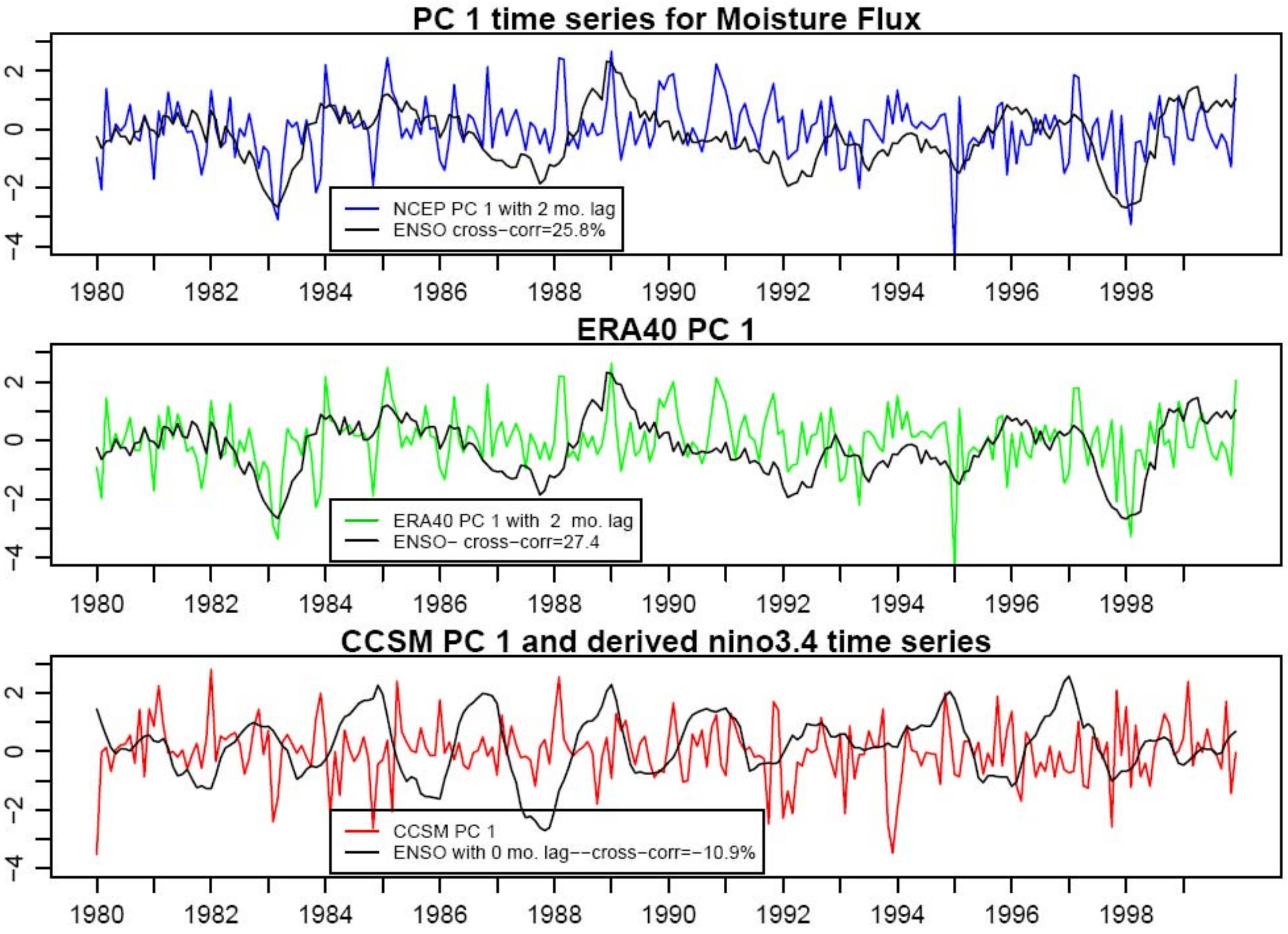

Figure 4.6: Comparison of moisture flux and ENSO variations for two reanalyses (NCEP and ERA40) and CCSM3.0.

Table 4.1: Comparison of zonal flow and fluxes at 130W with ENSO for reanalyses and CCSM3.0. Zonal flow and fluxes are characterized by PC1 of the monthly-mean time series over 1980 to 1999. ENSO is characterized by the Nino 3.4 SST time series. Numbers in the Table are correlation coefficients, followed in square brackets by the lag (positive lag means that the flow or flux lags behind ENSO) that gives the maximum correlation. The positive correlations for the flow/ENSO link indicate that zonal flows are stronger when Nino 3.4 SSTs are warmer. All flux correlations are shown as positive because the sign of the PC loading is arbitrary.

Field correlated with ENSO	NCEP	ERA40	CCSM3.0 (run e)
Zonal flow	0.523 [2 mo.]	0.538 [2 mo.]	0.155 [7 mo.] 0.328 [2 mo.][1]
Momentum flux	0.623 [2 mo.]	0.632 [2 mo]	0.286 [0 mo.]
Heat flux	0.514 [2 mo.]	0.531 [2 mo.]	0.211 [0 mo.][2]
Moisture flux	0.258 [2 mo.]	0.274 [2 mo.]	0.109 [0 mo.][2]

[1] The 2 mo. lag result is given for direct comparison with the NCEP and ERA40 results.

[2] Larger correlations occur at negative lags (ENSO lagging), which is physically unrealistic. Hence zero lag result is given instead.

4.2 Direct ENSO validation

We now consider the ENSO performance of other AOGCMs in the AR4/CMIP3 data base. The results reported below are from the paper by AchutaRao and Sperber (2006). We consider three different metrics for ENSO performance, the power spectrum, the temperature response pattern, and variability, comparing observed data with control-run model simulations. We do not examine model relationships between ENSO and fluxes.

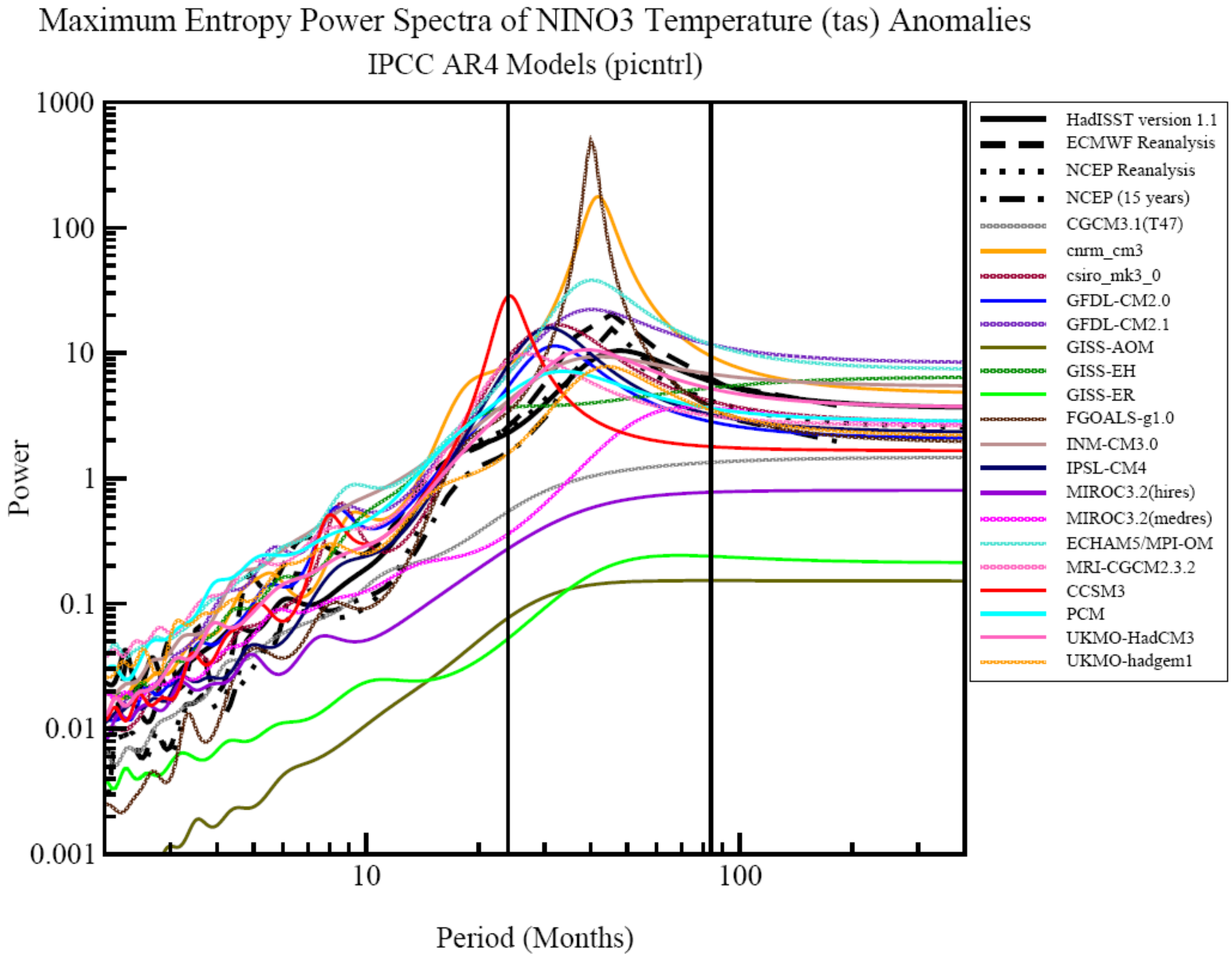

Figure 4.7: Power spectra for Nino3 SSTs for models in the AR4/CMIP3 data base.

Figure 4.7 shows the power spectra for these models. This is Fig. 2 from AchuaRao and Sperber (the same Figure is shown in the IPCC AR4 WG1 report, p. 624 – the version shown here is a higher quality plot obtained from Ken Sperber.) There a number of models where the ENSO spectrum is seriously deficient. The following models have virtually no SST periodicity (i.e., no identifiable ENSO signal in the appropriate part of the tropical Pacific) – CCCMA, GISS-EH, GISS-ER, GISS-AOM, MIROCmedres and MIROChires. The following models have periodicities that are far too strong (or at the wrong frequency) – CNRM, FGOALS and CCSM3.0. As the AchutaRao and Sperber analysis did not consider the ECHO-G and BCCR models, this leaves 10 models with acceptable SST spectra (see Table 4.2).

Table 4.2: ENSO performance metrics. Performance ranks only are given for RMSE and pattern correlation. The top 10 models are indicated by blue type. Cases where the model variability is too high (rather than too low) are shown as red in the last column.

Model	Power Spectrum	RMSE	Pattern Correl	N34 vblty index
BCCR	N.A.	N.A.	N.A.	N.A.
CCCMA	Poor	**8**	12	1811 [14]
CCSM3.0	Poor	**3**	**8**	183 [**7**]
CNRM	Poor	16	**3**	3055 [15]
CSIRO		14	14	325 [**9**]
ECHO-G	N.A.	N.A.	N.A.	N.A.
FGOALS	Poor	18	**10**	4116 [16]
GFDL2.0		**4**	**5**	204 [**8**]
GFDL2.1		17	**6**	1182 [12]
GISS-EH	Poor	13	17	4 [**1**]
GISS-ER	Poor	**5**	**9**	22456 [18]
INM		19	18	143 [**6**]
IPSL		**1**	**2**	125 [**5**]
MIROChires	Poor	12	13	5011 [17]
MIROCmedres	Poor	11	16	1075 [11]
MRI		**2**	**4**	466 [**10**]
ECHAM5		**10**	**1**	1741 [13]
PCM		**7**	11	20 [**3**]
HadCM3		**6**	**7**	63 [**4**]
HADGEM1		**9**	15	15 [**2**]
(GISS-AOM)	Poor	15	19	22456 [18]

The next ENSO metric we consider is the pattern of SST variations during an ENSO warm event (see AchutaRao and Sperber, Figures 3 and 4). Table 4.2 shows relative skill in the simulated pattern, quantified using both the RMS error and the observed versus model pattern correlations over 30S to 30N, 90 to 270W. The Table shows model rank, with the best models having the lowest RMSE values or the highest correlations. RMSE values range from 0.23 degC (IPSL) to 0.60 degC (INM), and pattern correlations range from 0.40 (GISS-AOM) to 0.86 (ECHAM5).

Our third metric is SST variability in the Nino 3.4 region. These results are given in AchutaRao and Sperber's Table 2. Since both too high and too low variability (relative to observations) are undesirable, we use the following index …

$$I = 10000\,[0.5\,(s/z + z/s) - 1]$$

where s and z are the observed (HadISST v1.1) and model standard deviations. Small values of I indicate a better fit, and I=0 for a perfect fit. Cases where the model variability is too high (rather than too low) are shown as red in the Table.

Overall, the best models are IPSL, MRI, HadCM3 and GFDL2.0. CCSM3.0 also performs well except for its strong 2-year periodicity. Note that MRI is flux adjusted, and its performance may be partly an artifact of the flux adjustment. The worst models are FGOALS, MIROChires, INM and HadGEM1.

The version of CCSM used in the above analyses (3.0) is the version archived in the AR4/CMIP3 data base. This is now a few years out of date (as is the case for most of the models in this data base) and substantial improvements have been made since then to CCSM's ENSO performance. Figure 4.8 (from Jerry Meehl, NCAR, pers. comm..) shows this. It is clear that the spectrum has improved markedly (there is no longer the pronounced 2-year cycle). The variability, however, is still somewhat larger than the observed variability.

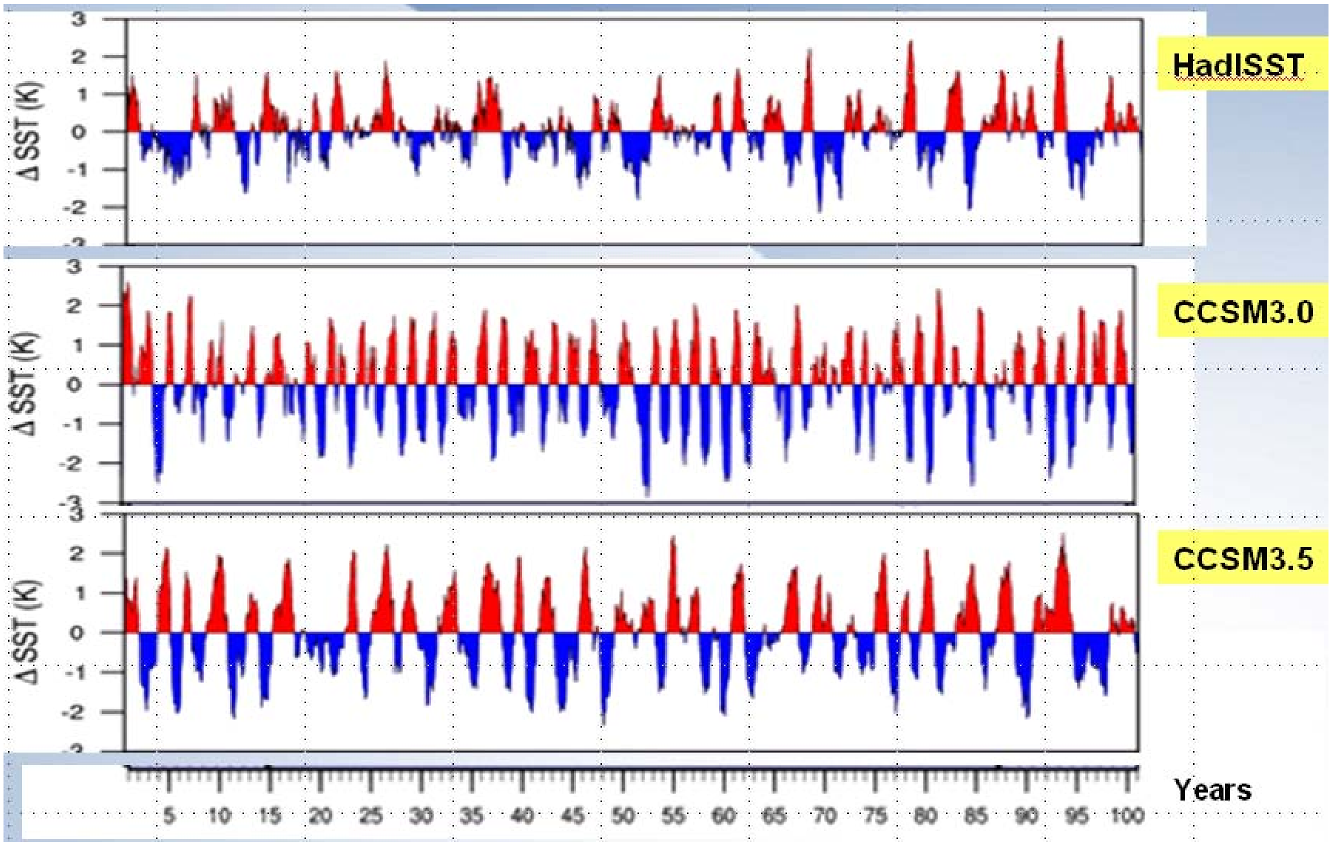

Figure 4.8: Nino3 SST simulations comparing observed changes with two versions of CCSM. Note that the time scale is arbitrary.

4.3 Summary to date

At this stage, the models may be separated into three groups, those that are clearly superior, those that are clearly inferior, and those that are in between.

Superior models are CCSM3.0, GFDL2.0, GFDL2.1, IPSL, MIROCmedres and HadCM3. Inferior models that have levels of performance that make them unsuitable as RCM drivers are CNRM, FGOALS, GISS-EH, GISS-ER, INM and PCM. Models that are in between are CCCMA, ECHO-G and MRI (which are mainly suspect because they are flux adjusted) BCCR, CSIRO, MIROChires, ECHAM5 and HadGEM1.

5.0 Flux Assessments for CCSM3.0

Fluxes across the western boundary are determined by the westerly wind strength and the prevailing field for the flux variable. For momentum flux, therefore, the latitude-height flux pattern will closely mirror the pattern for the wind (u). This will also be the case for heat flux, as noted above. The situation for moisture fluxes is quite different. Here, both moisture content (specific humidity) and wind speed vary substantially both latitudinally and vertically, so both will affect the flux pattern. It is the spatial moisture variations that have the dominant effect. The key determinants of fluxes, therefore, are the westerly wind and specific humidity latitude-height patterns over the boundary at 130W (= 230E).

5.1 Variable profiles

We begin with an analysis of the wind speed, temperature and moisture fields over the western boundary, comparing CCSM3.0 data with reanalyses.

Figures 5.1, 5.2 and 5.3 compare the westerly wind speed fields, the temperature fields and the specific humidity fields for CCSM3.0 with ERA40 for months that are central to (and typical of) the winter, spring, summer and fall seasons. For the wind fields (Figure 5.1), CCSM has stronger jet maxima, especially in the winter/spring half of the year (see also below). The implications of this are not immediately clear. As noted above, moisture fluxes are probably more important than other fluxes, and the positions of the anomalous jet speed maxima are too high to affect moisture fluxes significantly. Below about 700 hPa, the CCSM and ERA wind fields are quite similar. Note that the mean wind direction south of 30N is easterly in the lower troposphere in both the model and the observations.

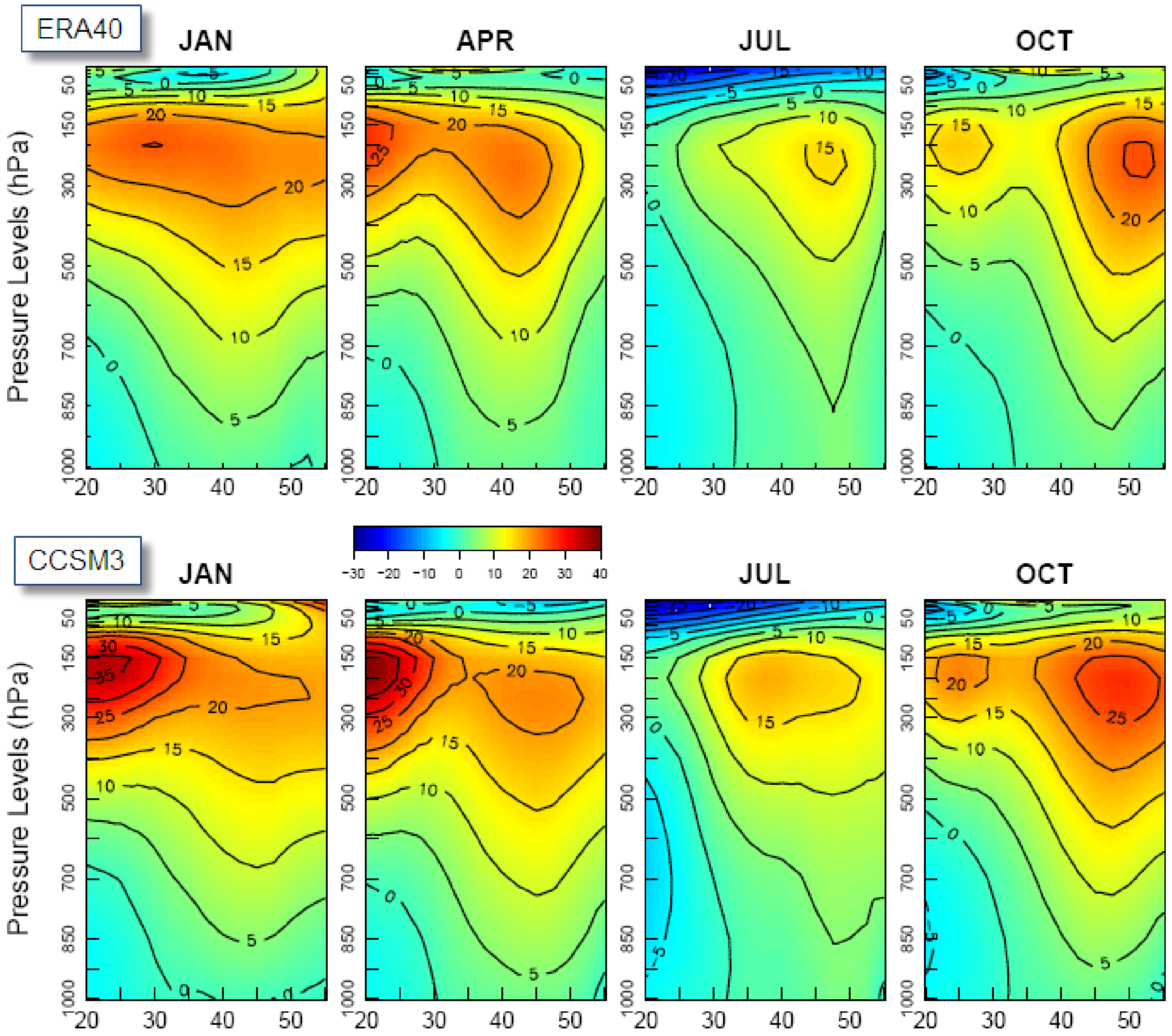

Figure 5.1: Vertical profiles for westerly wind speeds (m/s) along the western boundary (130W)

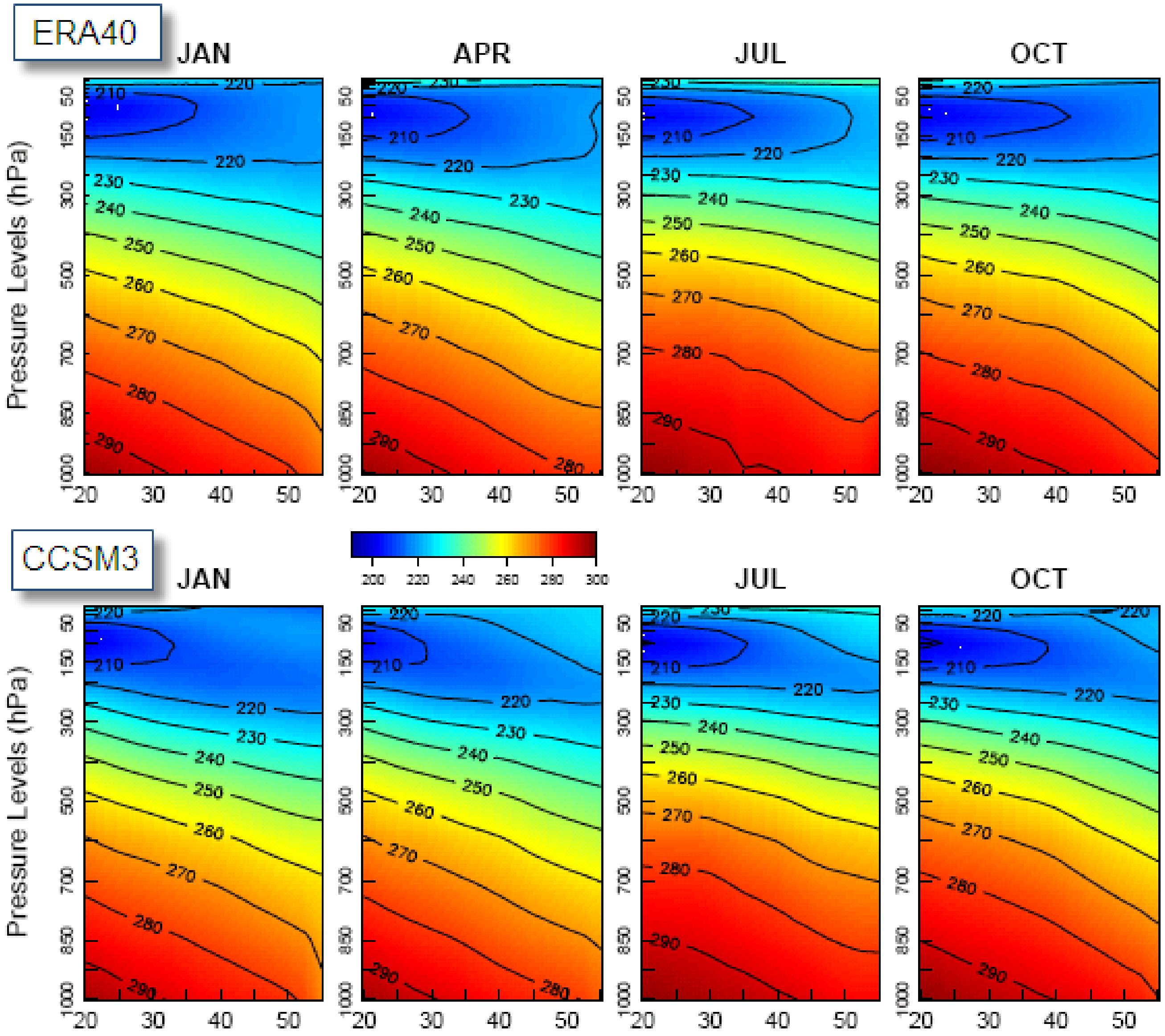

Figure 5.2: Vertical profiles for temperature (K) along the western boundary at 130W

Temperature fields (Figure 5.2) are very similar in CCSM and the observations. The differences are so small that they can have only negligible effects on heat fluxes.

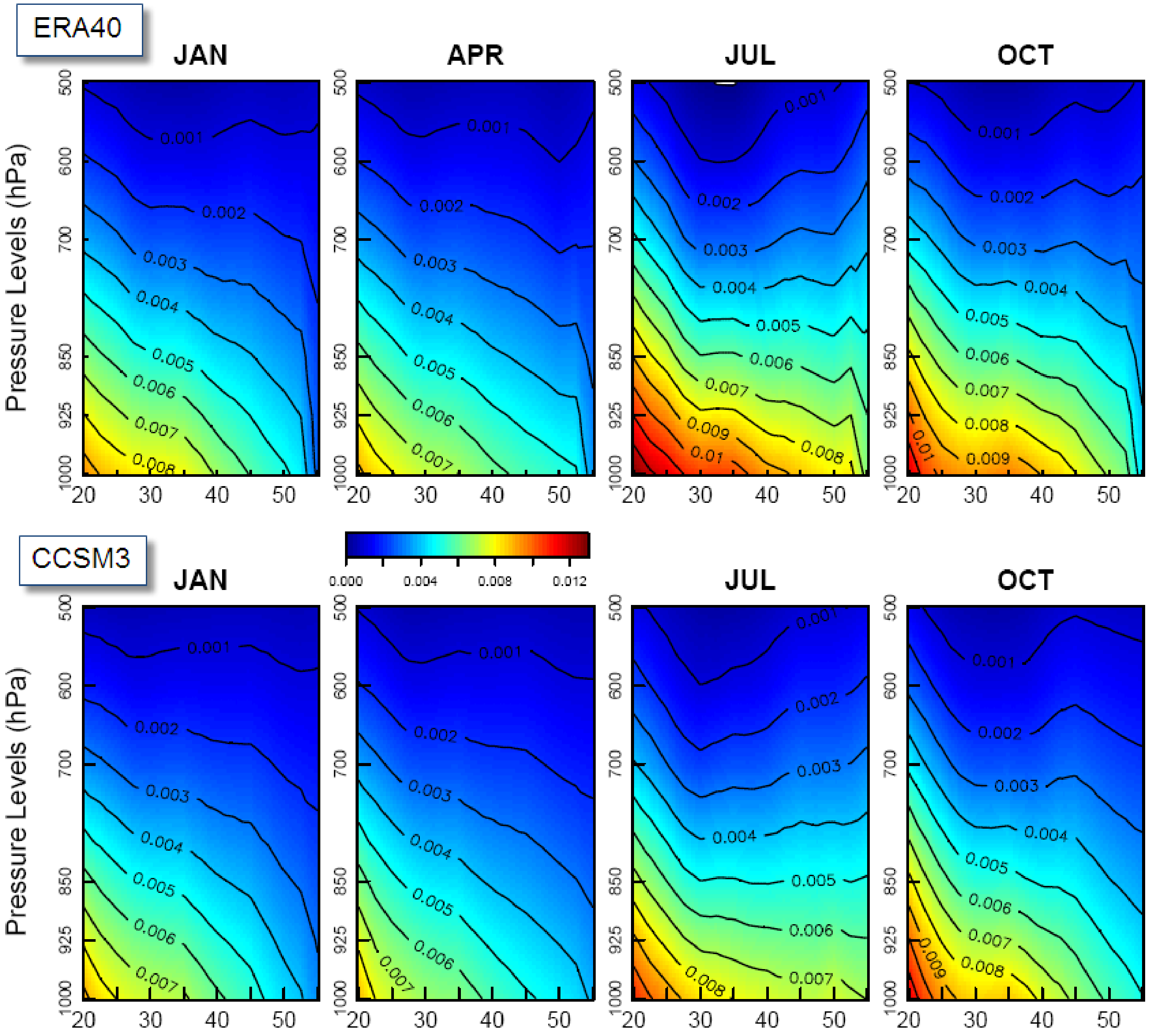

Figure 5.3: Vertical profiles for moisture content (specific humidity, gm/gm) along the western boundary at 130W

Moisture fields (Figure 5.3) show larger differences. Near the surface, model specific humidities are around 10% less than observations in January, and around 20% less in July. It is the January differences that are most important, since this is the wet season. As the westerly wind speeds in the model and observations are very similar in the lower troposphere, one might expect moisture fluxes in the model to be less than observed (this is, however, not the case – see below).

Another test for model simulations of fluxes is to examine how well models simulate the observed seasonal cycles of jet latitude and jet speed. Note that, at certain times of the year, there are two distinct jets over the Pacific, both a subtropical jet at lower latitudes and a polar jet at higher latitudes (see the April and October plots in Figure 5.1). Figures 5.4 and 5.5 compare observed and model (CCSM3.0) seasonal cycles of jet latitude and jet strength. The model simulates the jet locations very well over the seasonal cycle (Figure 5.4), but is less good at simulating the strength of the subtropical jet (Figure 5.5). CCSM3.0 produces considerably higher jet speeds than observed, as already noted, especially in December through April.

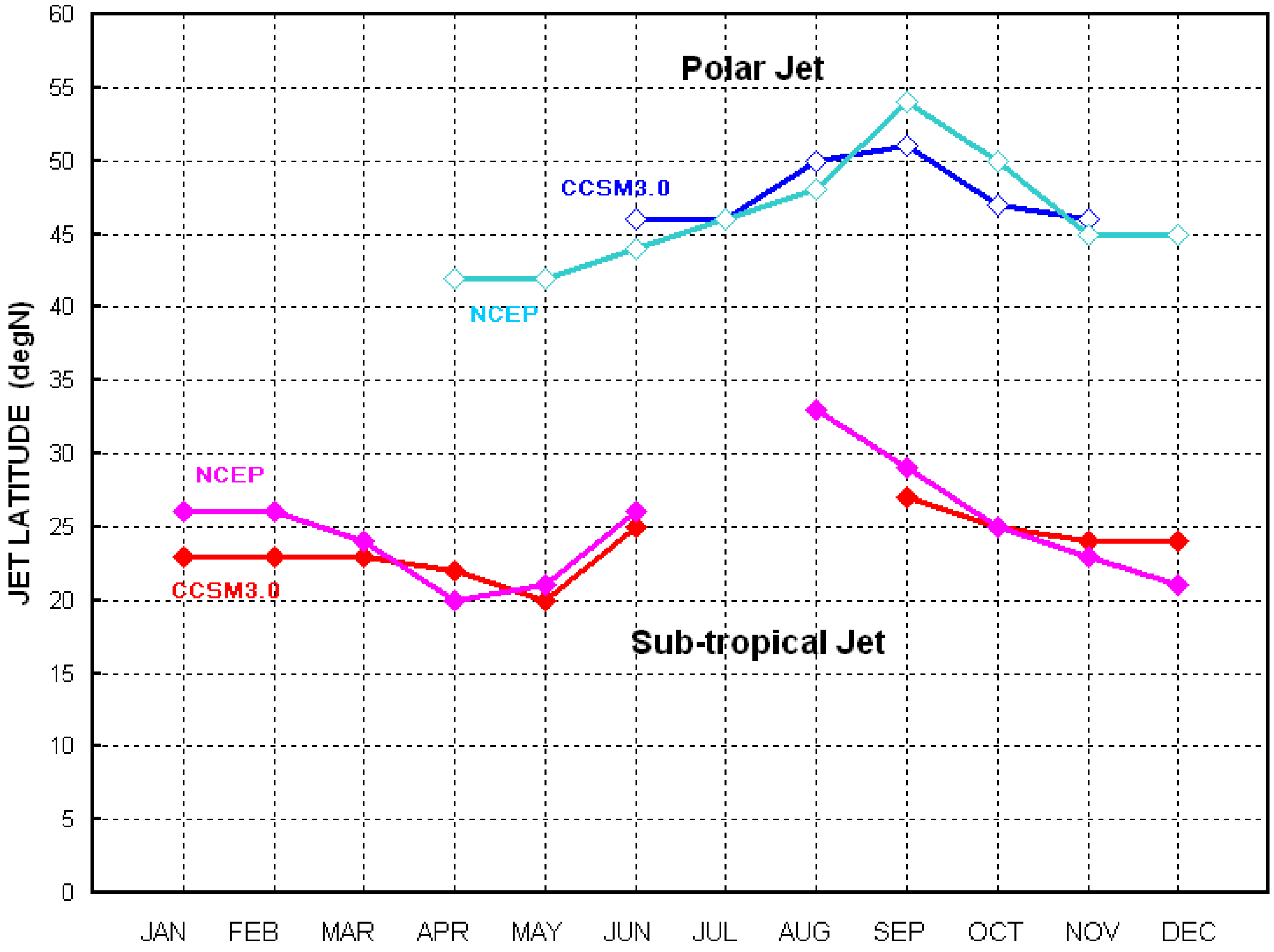

Figure 5.4: Seasonal movements of jet latitudes in the model (CCSM3.0) and reanalysis (NCEP). Given the difficulty in precisely defining jet location, the agreement is very good.

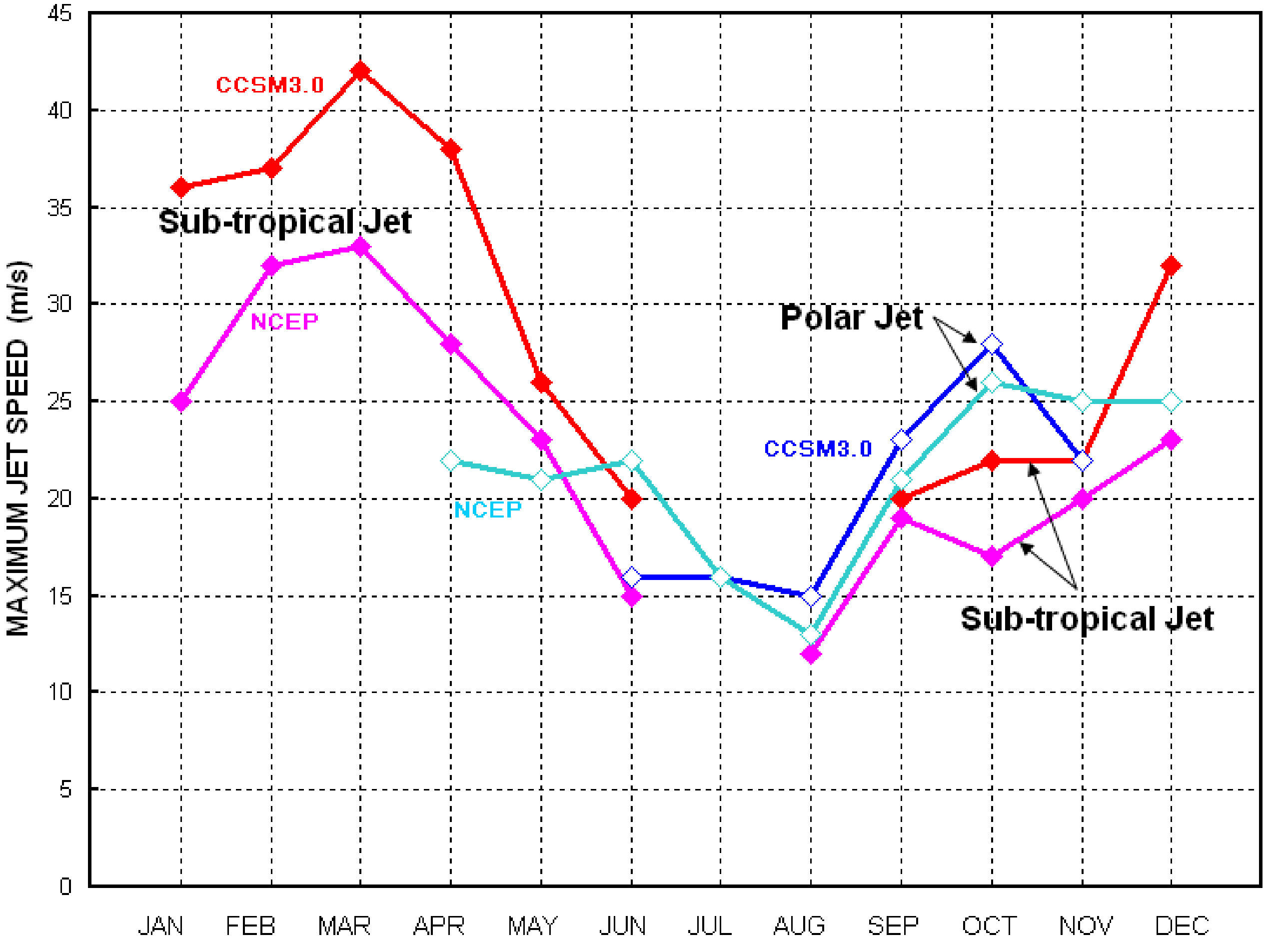

Figure 5.5: Seasonal changes in jet strengths in the model (CCSM3.0) and reanalysis (NCEP). The model seriously over-estimates the strength of the sub-tropical jet.

5.2 Flux profiles

We noted above that true fluxes can only be calculated using high temporal resolution (6-hourly) data, and that, of the models in the AR4/CMIP3 data base, such high temporal resolution data are available only for CCSM3.0. We therefore introduced the concept of pseudo fluxes, which can be calculated using monthly data. Real fluxes (for variable X) are defined by <uX>, whereas pseudo fluxes are defined by <u><X>, where <...> denotes a time average.

In this section we consider both real and pseudo fluxes for CCSM3.0 and compare these with real and pseudo fluxes based on the ERA40 reanalysis data. Figure 5.6 compares both real and pseudo momentum fluxes for CCSM3.0 with those for ERA40 data.

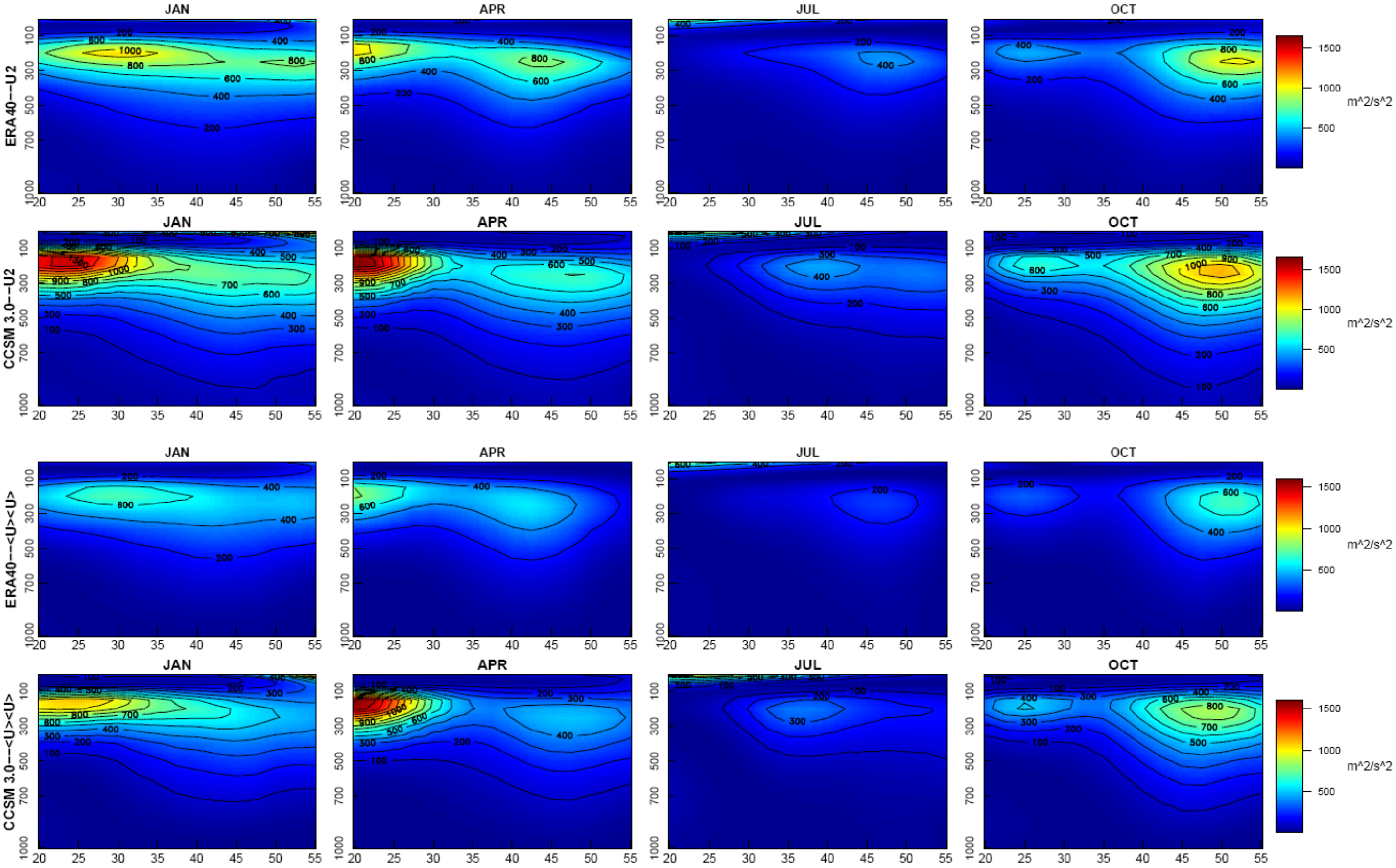

Figure 5.6: Vertical profiles of momentum fluxes across the western boundary at 130W. Results for both true fluxes (<uu>) and pseudo fluxes (<u><u>) are shown, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are m^2/s^2.)

These plots are similar to those for the westerly wind field (Figure 5.1). The top two rows show the same relative results as Figure 5.1, with fluxes in the region of the subtropical jet in January and April being appreciably stronger in CCSM3.0 than in the observations. In the lower troposphere, however, model and observed fluxes are similar. The pseudo flux patterns (bottom two rows) are very similar to the true flux patterns, but with weaker maxima. The pseudo flux results for the model show the same relative errors as for the true fluxes. It is clear that a comparison of pseudo momentum fluxes leads to the same conclusions as a comparison of real momentum fluxes.

Figure 5.7 compares both real and pseudo heat fluxes for CCSM3.0 with those for ERA40 data.

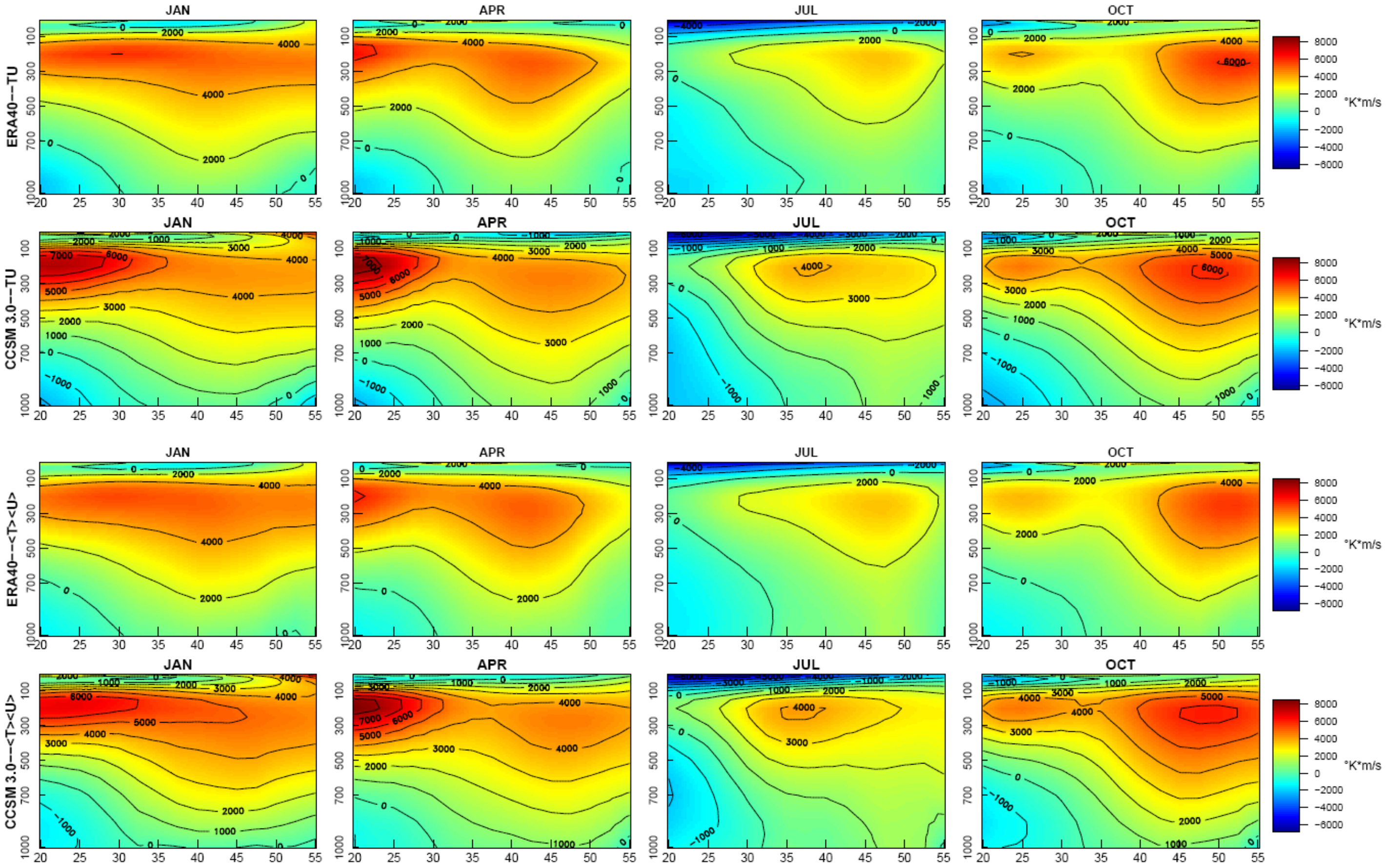

Figure 5.7: Vertical profiles of heat fluxes across the western boundary at 130W. Results for both true fluxes (<uT>) and pseudo fluxes (<u><T>) are shown, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are K m/s.)

Consider the real fluxes first. As anticipated, the latitude-height patterns of the heat fluxes are controlled mainly by the westerly wind structure, with maximum fluxes at the positions of the jet maxima. The seasonal variations also follow the seasonal variations in the westerly wind speed. There is a region of easterly flux south of 30N in January and April, extending a little further north in July and October. At 20N this runs up to around 700 hPa in January, April and October, and to around 400 hPa in July. Although the heat fluxes in the upper troposphere are larger in the model than in the observations, model and observed heat fluxes below around 500 hPa are virtually identical. (Note that some plots show contour lines at 1000 K m/s, while others have a 2000 K m/s interval.)

For pseudo fluxes the relative results, model versus observed, are the same as for real fluxes, so, again, the use of pseudo fluxes will lead to the same conclusions as for real fluxes.

Figure 5.8 shows results for moisture fluxes.

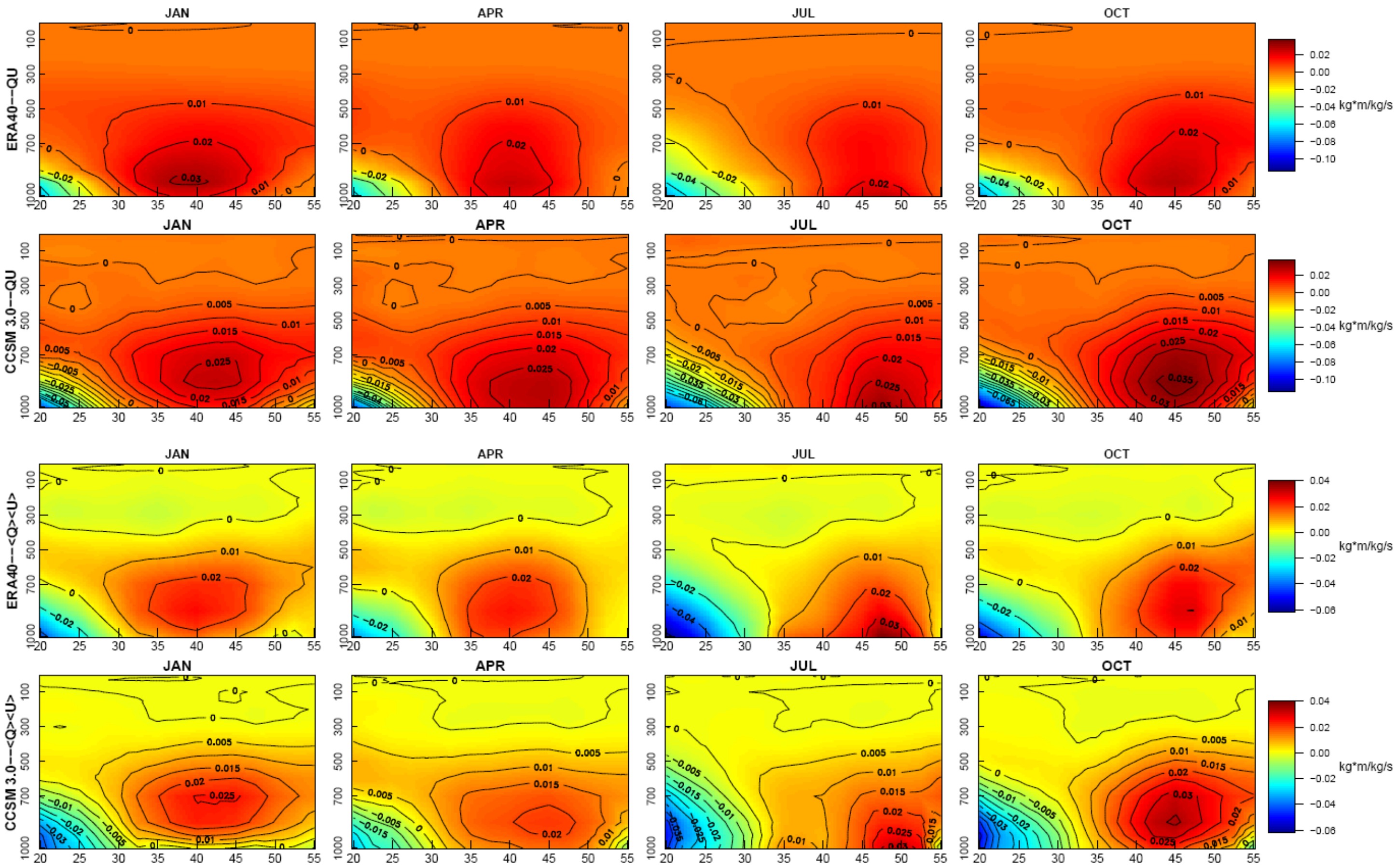

Figure 5.8: Vertical profiles of moisture fluxes across the western boundary at 130W. Results for both true fluxes (<uq>) and pseudo fluxes (<u><q>) are shown, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.)

As anticipated, the latitude-height patterns reflect the influences of both the wind speed and specific humidity patterns. Largest moisture fluxes occur in the lower troposphere where specific humidity levels are greatest. South of around 30N and up to around 700 hPa the fluxes are easterly due to dominantly easterly winds in these regions. There are clear flux maxima are around 850 hPa in January and October, and nearer to the surface in April and July. The maxima occur around 40 to 45N in January and April, and slightly further north in July and October. These are latitude ranges that correspond to the latitudes of the northern half of California. These features are common to both the model and the observations.

The seasonal cycles of moisture fluxes differ between model and observations. Both the observations and the model show a maximum in October (although the seasonal cycle range is quite small, much less than for precipitation), but the model shows a clear minimum in April. Model fluxes are noticeably less than the observed (ERA40) fluxes – for pseudo fluxes the deficits range from around 40% in April to 10% less in October (24% less in January).To our knowledge these moisture flux maxima have not previously been identified. As a new (and somewhat unexpected) feature it is a reassuring measure of model skill that they are well simulated in the model.

A key issue is whether model/observed moisture flux differences can help to explain the fact that all AOGCMs over-estimate precipitation in the western USA. We showed this problem in Figure 2.1 using model average results, noting (see Table 2.1) that it was a feature common to all models. Figure 5.9 shows that it is a significant problem in CCSM3.0 in particular (see also Figure 2.2).

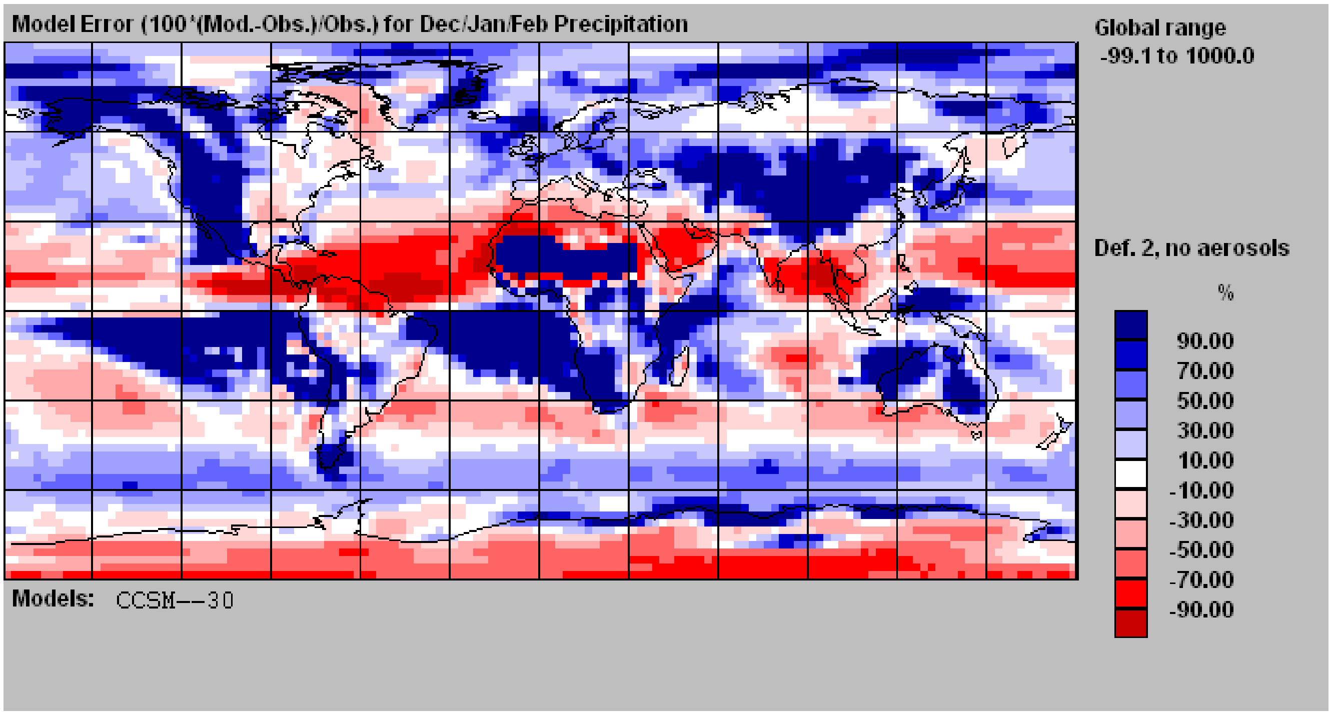

Figure 5.9: Bias in the CCSM3.0 model simulation of DJF precipitation. The map shows the model-mean percentage error. Blue areas show where the model simulation is too wet, red areas are where the model simulation is too dry.

It is clear that moisture fluxes biases in the model cannot explain these precipitation errors. In the lower troposphere, moisture fluxes in the model are slightly larger than in the observed data in April, July and October. In January, however, model and observed flux magnitudes are very similar. These differences are small compared with the precipitation errors, which are up to over 100% (i.e., the model gives DJF values that are more than double those observed) over much of the western USA. This conclusion is consistent with the less comprehensive analysis of Caldwell et al. (2009).

Figure 5.8 shows that pseudo fluxes for moisture are very similar in both patterns and magnitude to the real fluxes. Once again we can conclude that the same inferences would be drawn from pseudo fluxes as from real fluxes.

6.0 Flux Assessments for Other Models

In this section we consider four other AOGCMs from the AR4/CMIP3 data base, two that have been judged on the criteria already considered to be superior models (GFDL2.1 and MIROCmedres) and two that hav been judged to be inferior models, unsuitable as an RCM driver (PCM and GISS-EH). As noted previously, we only have monthly data for these models, so we must use pseudo fluxes for our flux assessment. We have already shown using CCSM3.0 data that pseudo fluxes are a perfectly adequate measure for flux assessment.

6.1 Presentation of Results

Figures 6.1 to 6.5 compare results for momentum fluxes for these four models with those for CCSM3.0 and the ERA40 reanalysis. For the models we show two or three realizations, i.e., runs with the same forcing but with different initial states.

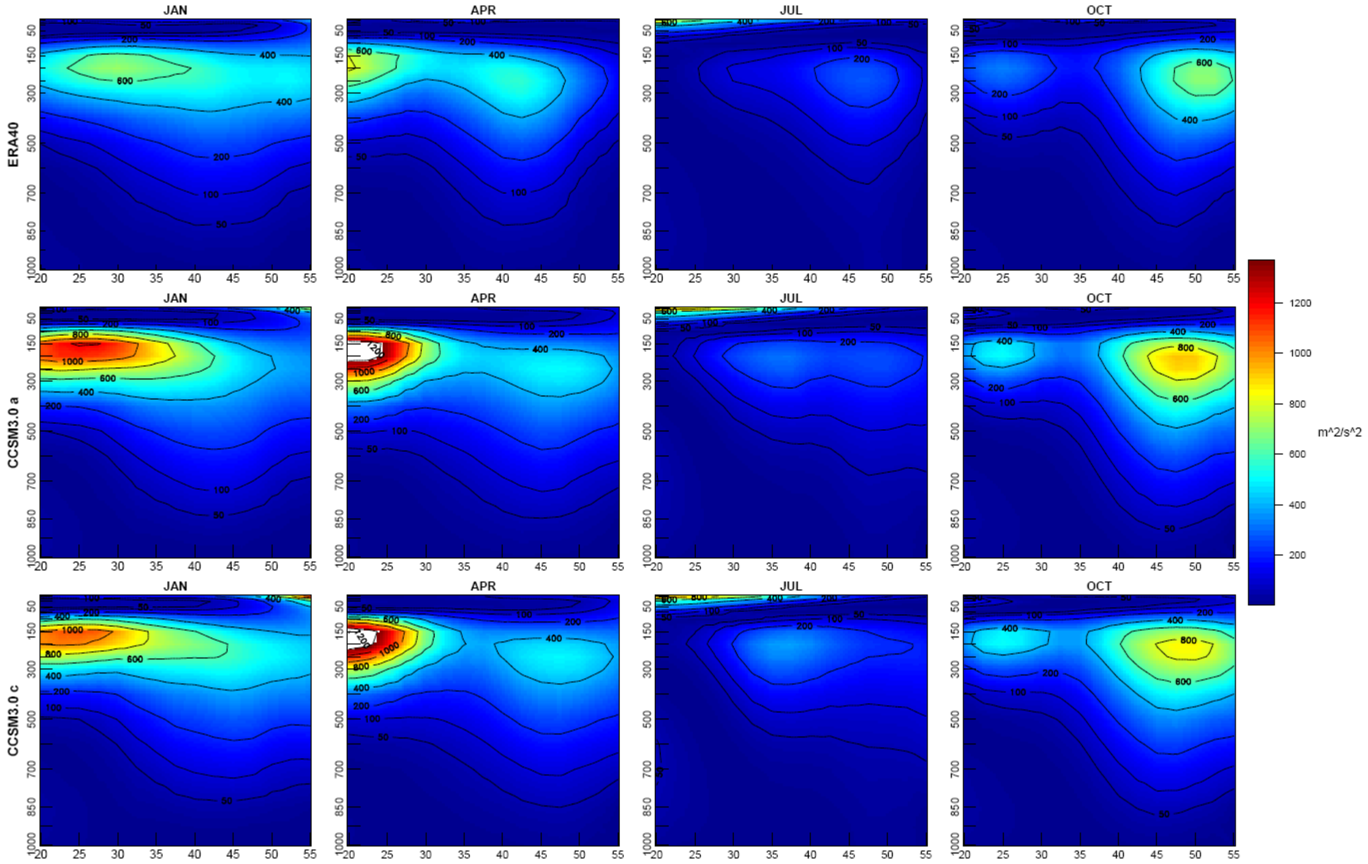

Figure 6.1: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing CCSM3.0 model results with observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

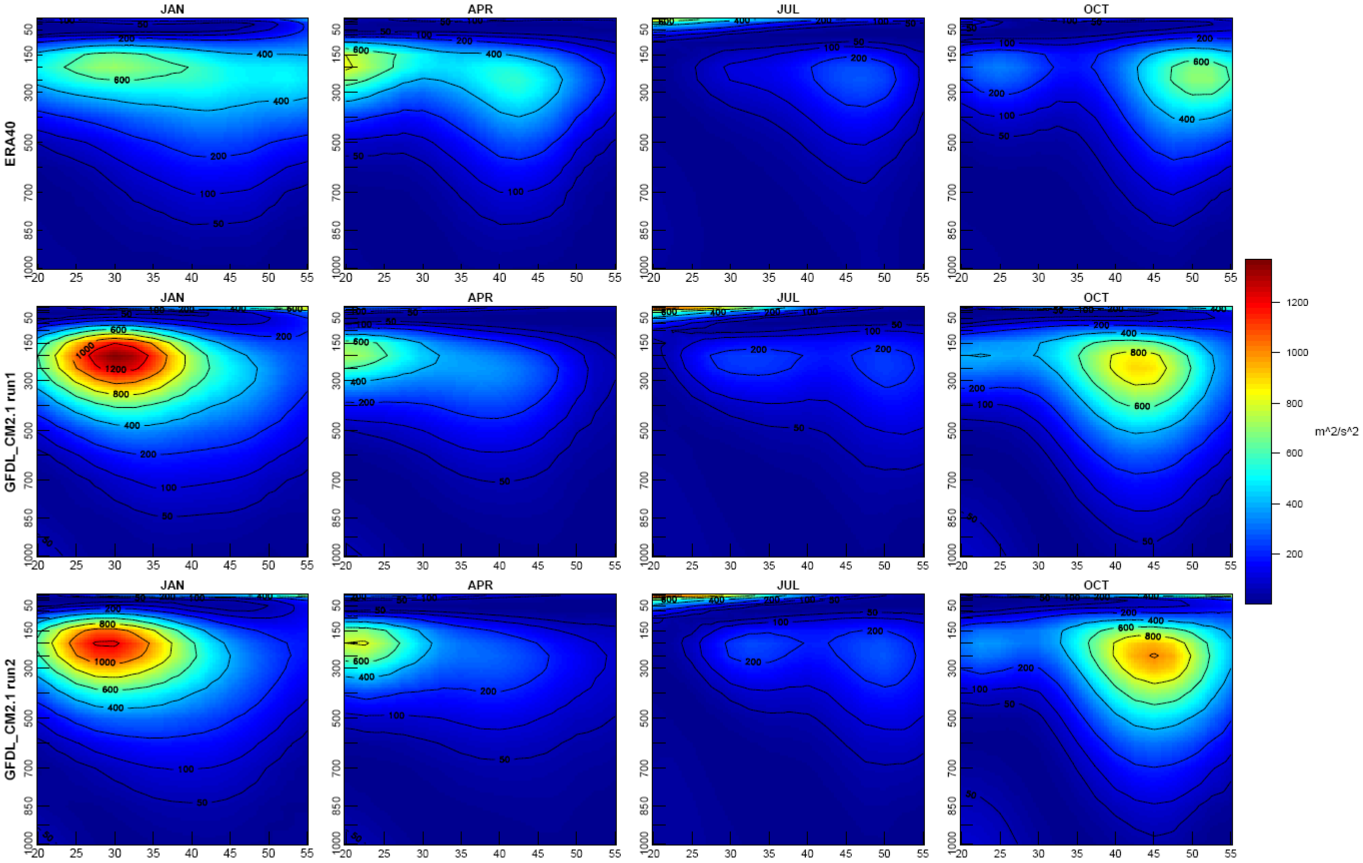

Figure 6.2: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (GFDL2.1) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

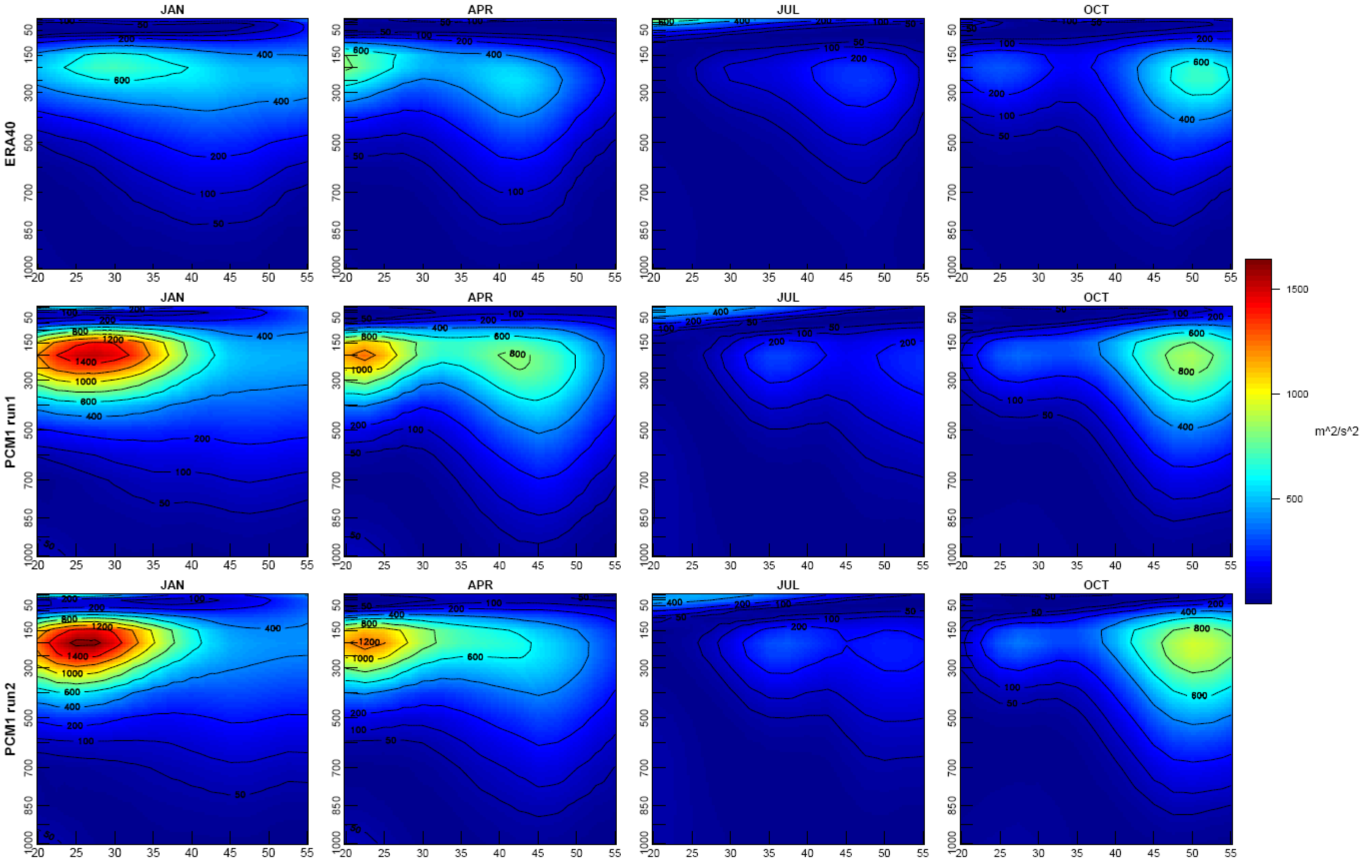

Figure 6.3: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (PCM) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

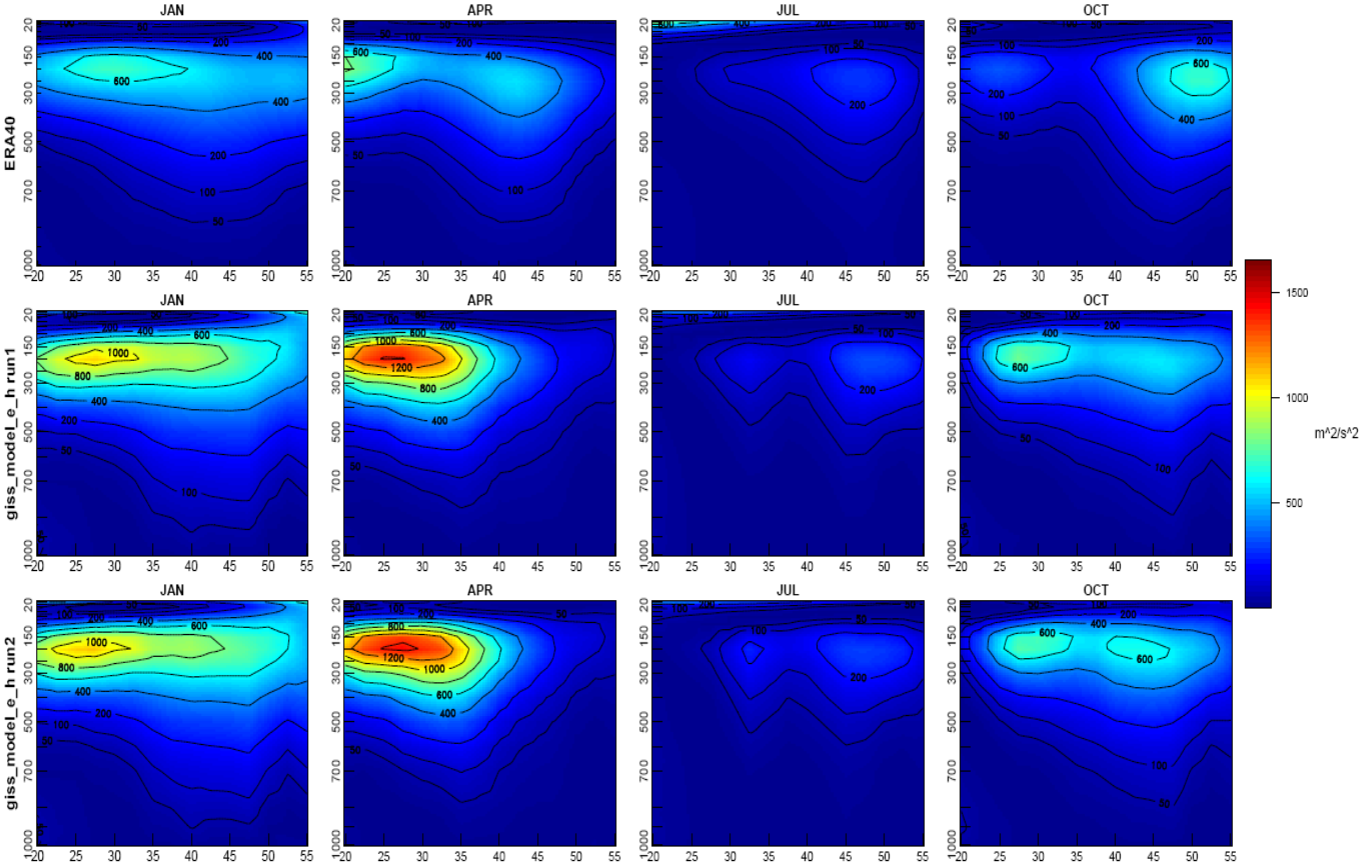

Figure 6.4: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (GISS-EH) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Two runs (using different initializations) are shown for the model – bottom two rows.

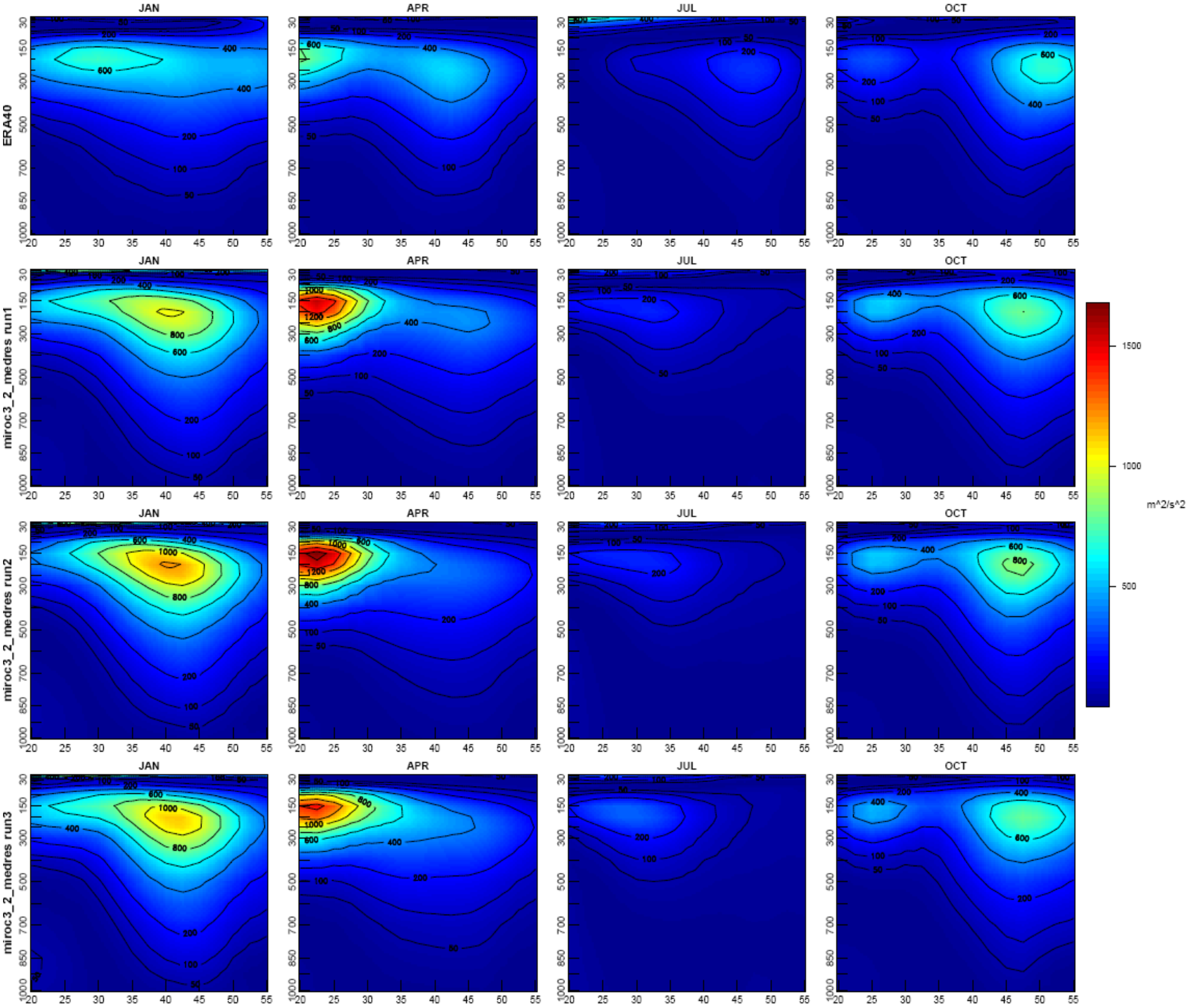

Figure 6.5: Vertical profiles of pseudo momentum fluxes across the western boundary at 130W, comparing model (MIROCmedres) and observed (ERA40 reanalysis – top row) results. (Units are m^2/s^2.). Three runs (using different initializations) are shown for the model – bottom three rows.

Figures 6.6 to 6.10 show results for the pseudo heat fluxes.

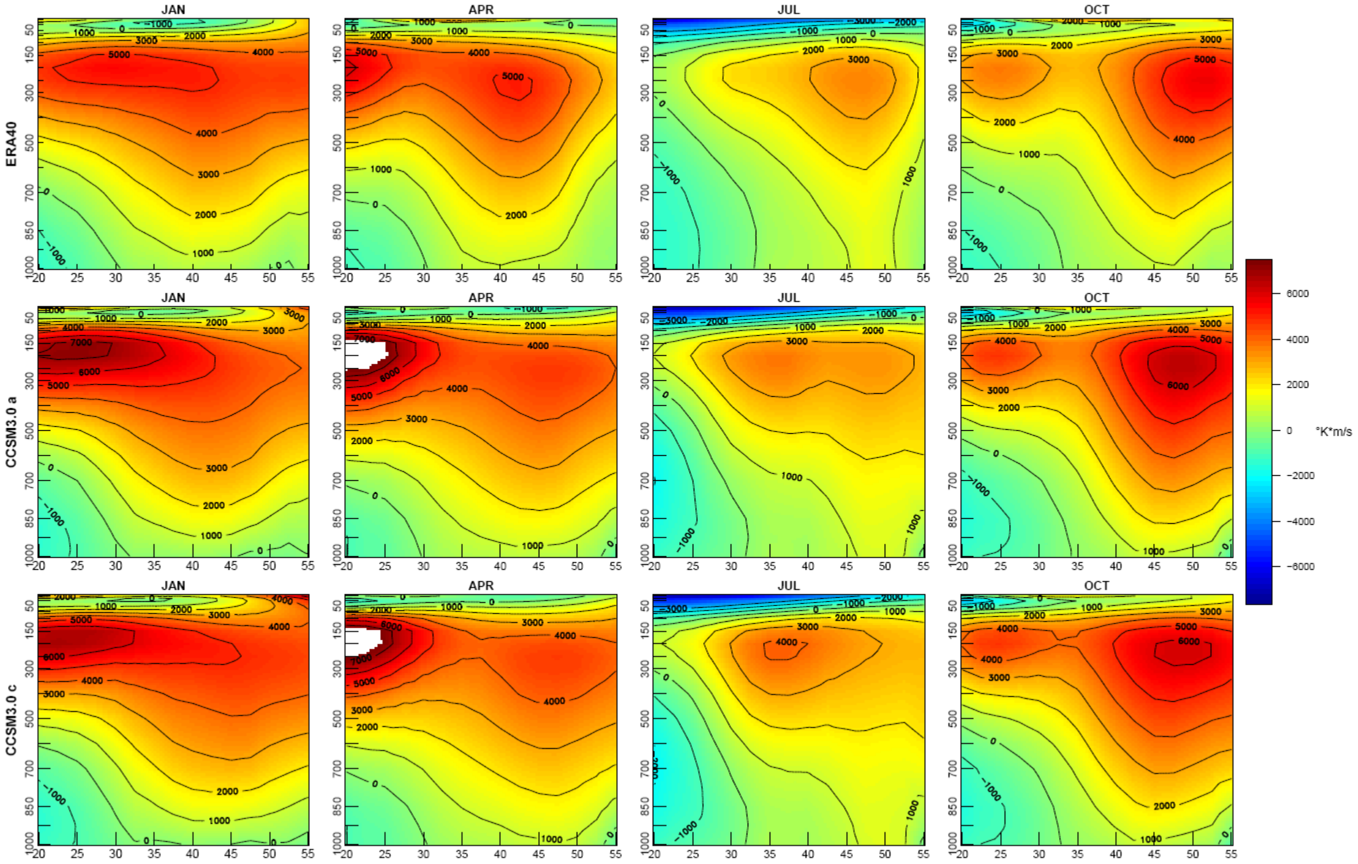

Figure 6.6: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

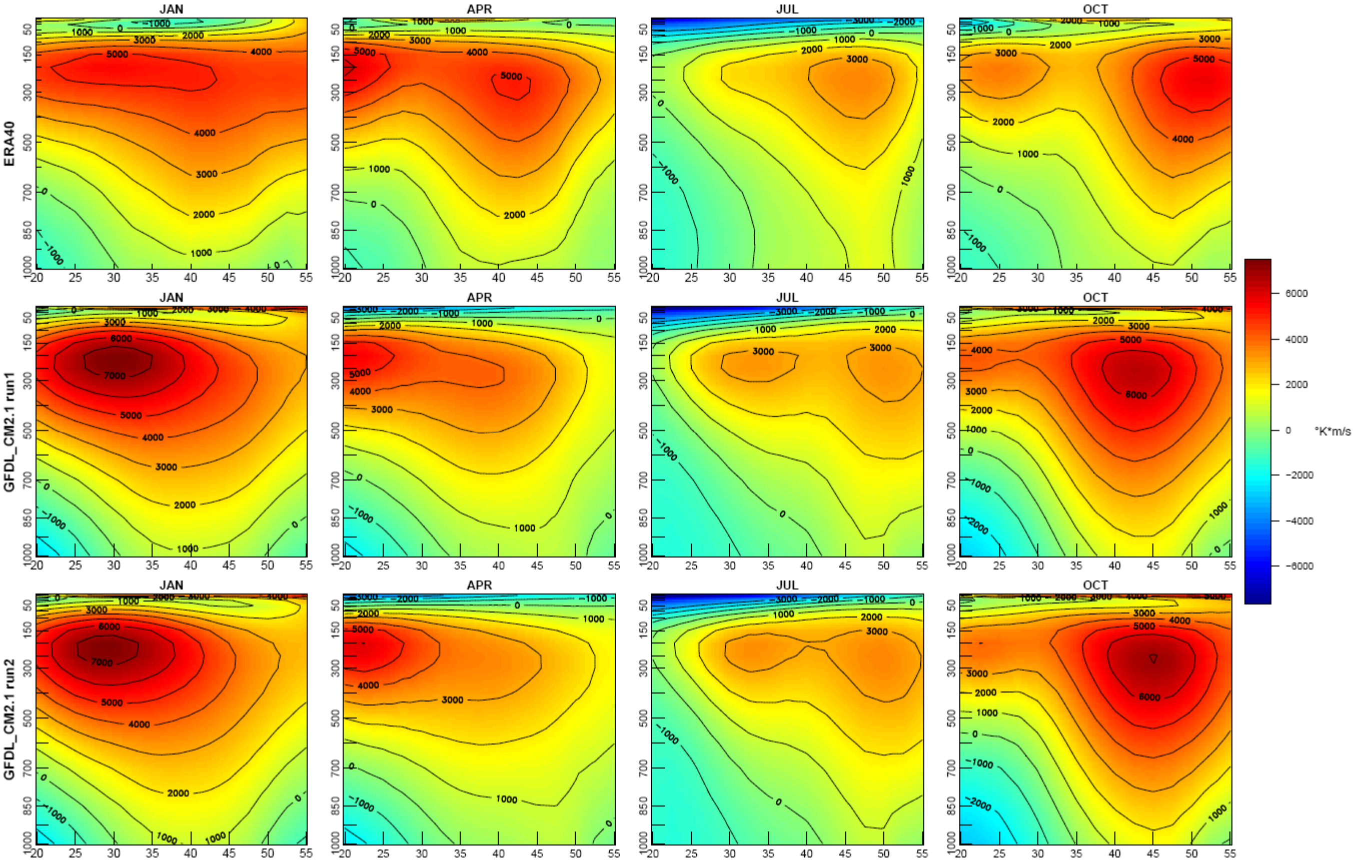

Figure 6.7: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (GFDL2.1) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

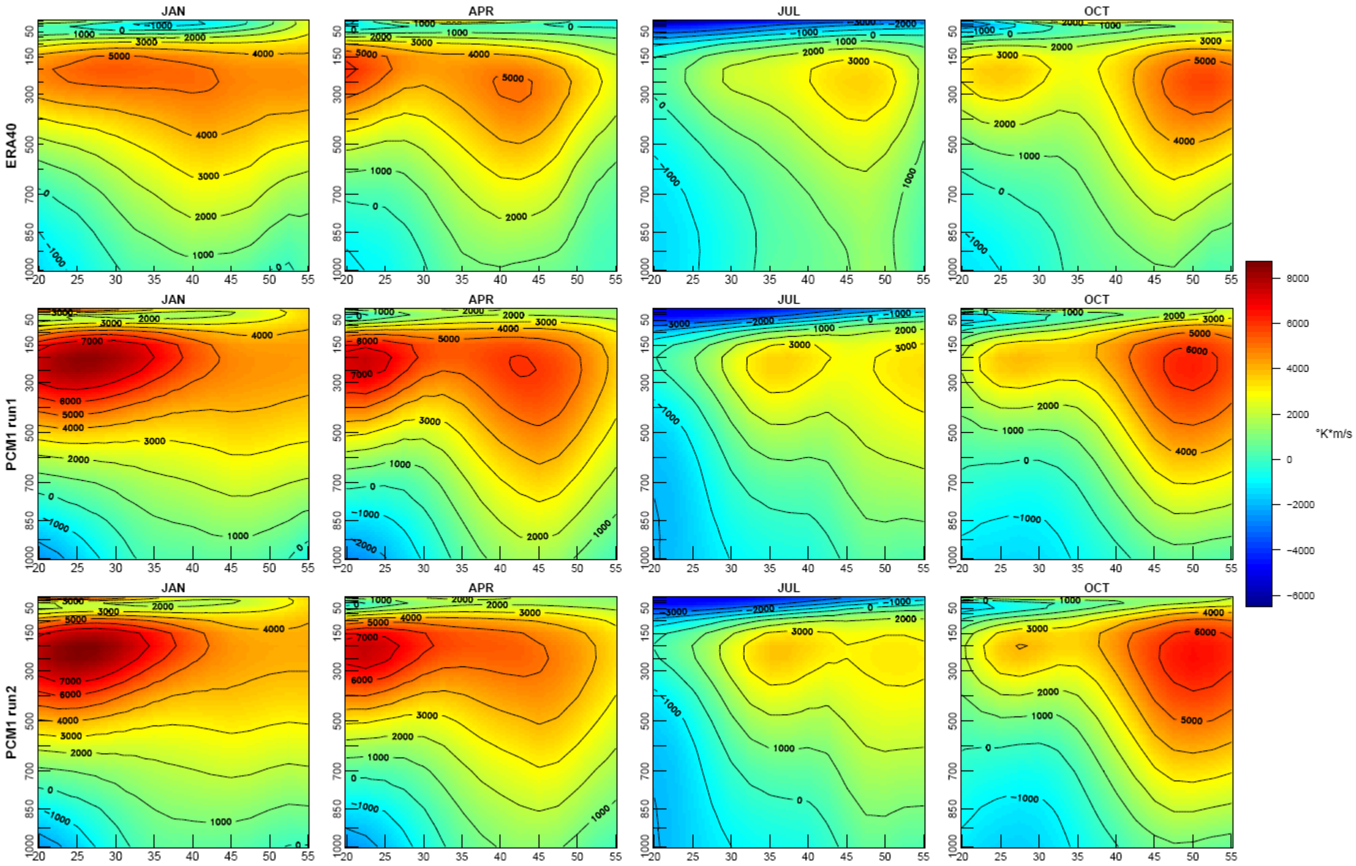

Figure 6.8: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (PCM) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

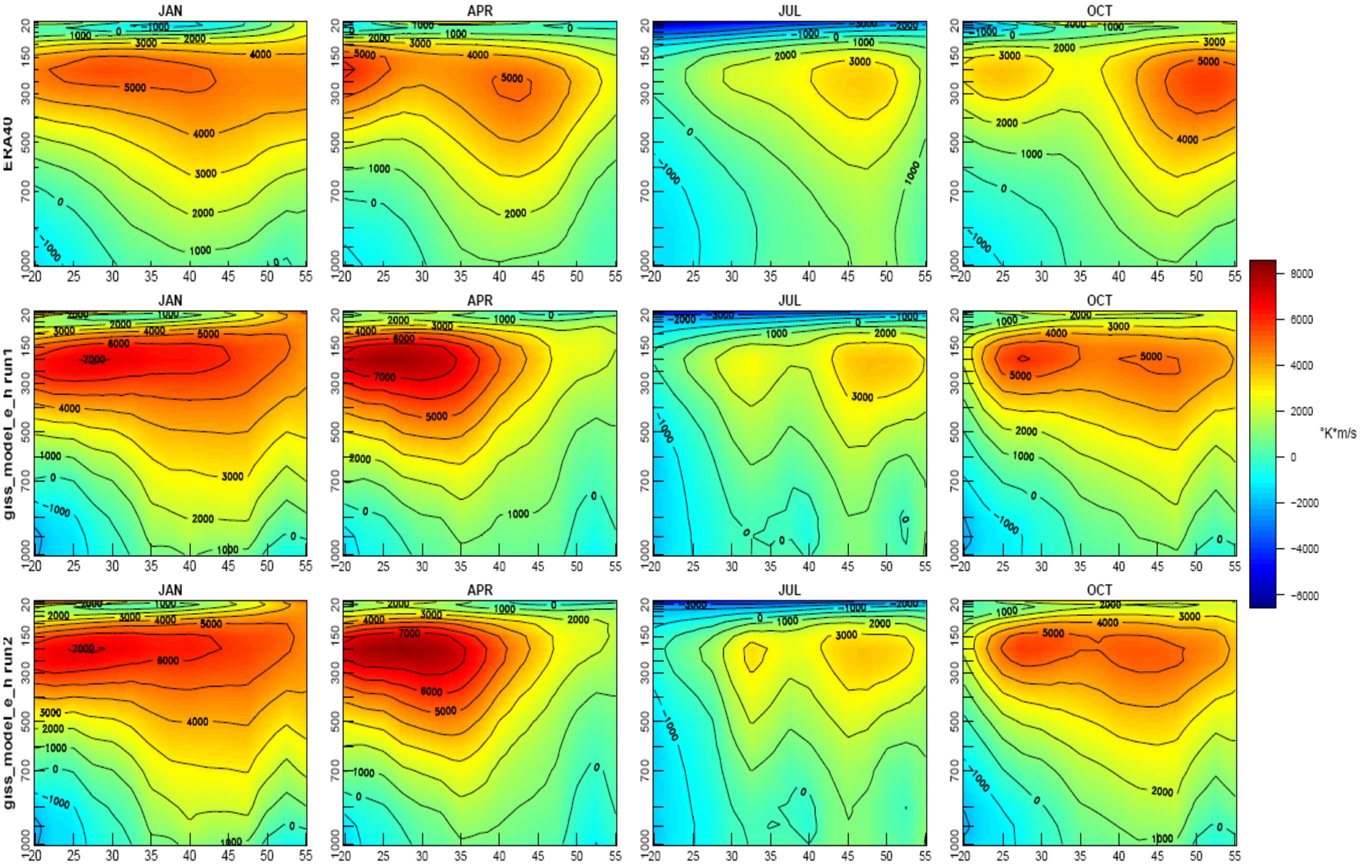

Figure 6.9: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (GISS-EH) and observed (ERA40 reanalysis) results. (Units are K m/s.). Two runs (using different initializations) are shown for the model.

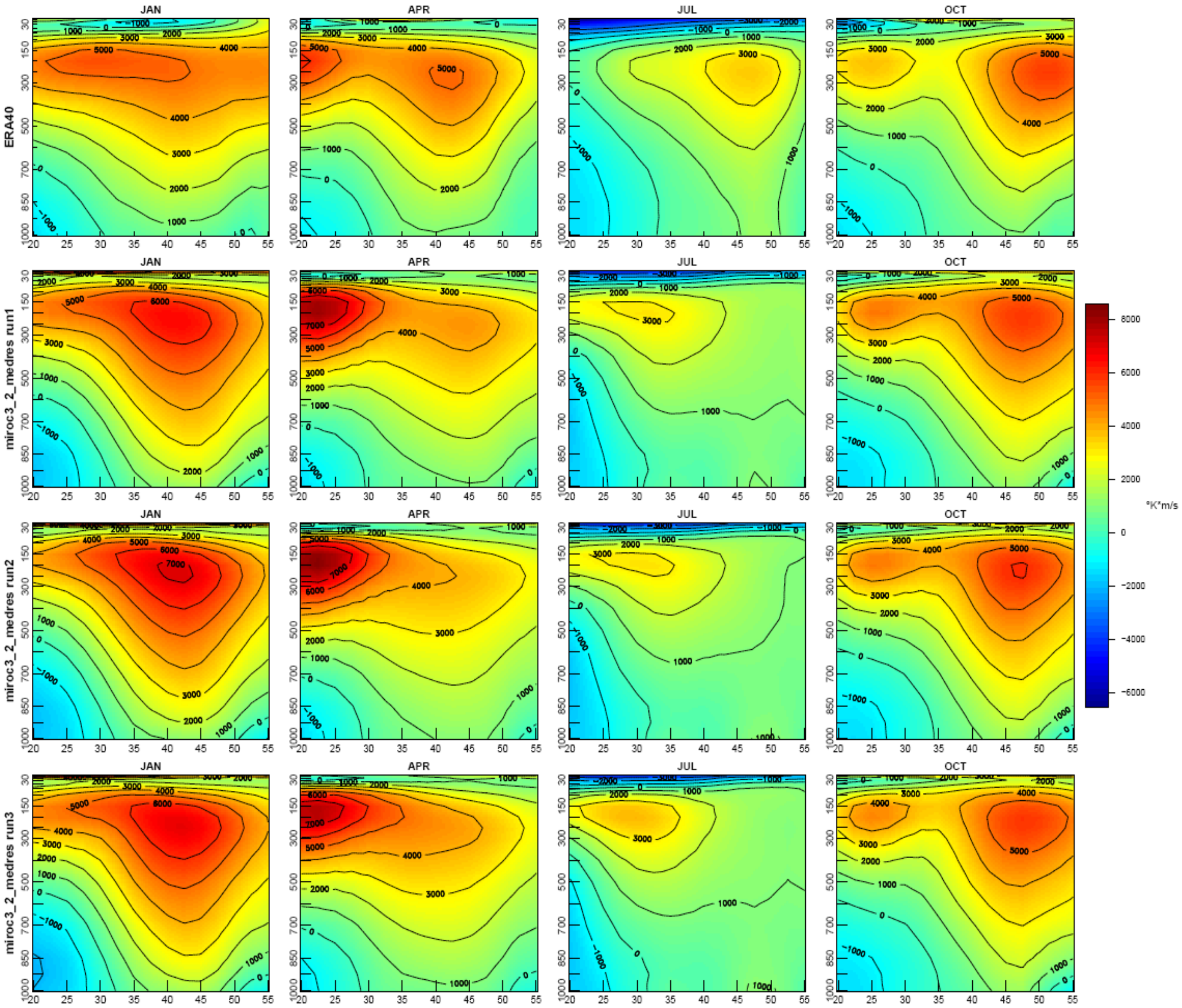

Figure 6.10: Vertical profiles of pseudo heat fluxes across the western boundary at 130W, comparing model (MIROCmedres) and observed (ERA40 reanalysis) results. (Units are K m/s.). Three runs (using different initializations) are shown for the model.

Moisture flux results are shown in Figures 6.11 to 6.15.

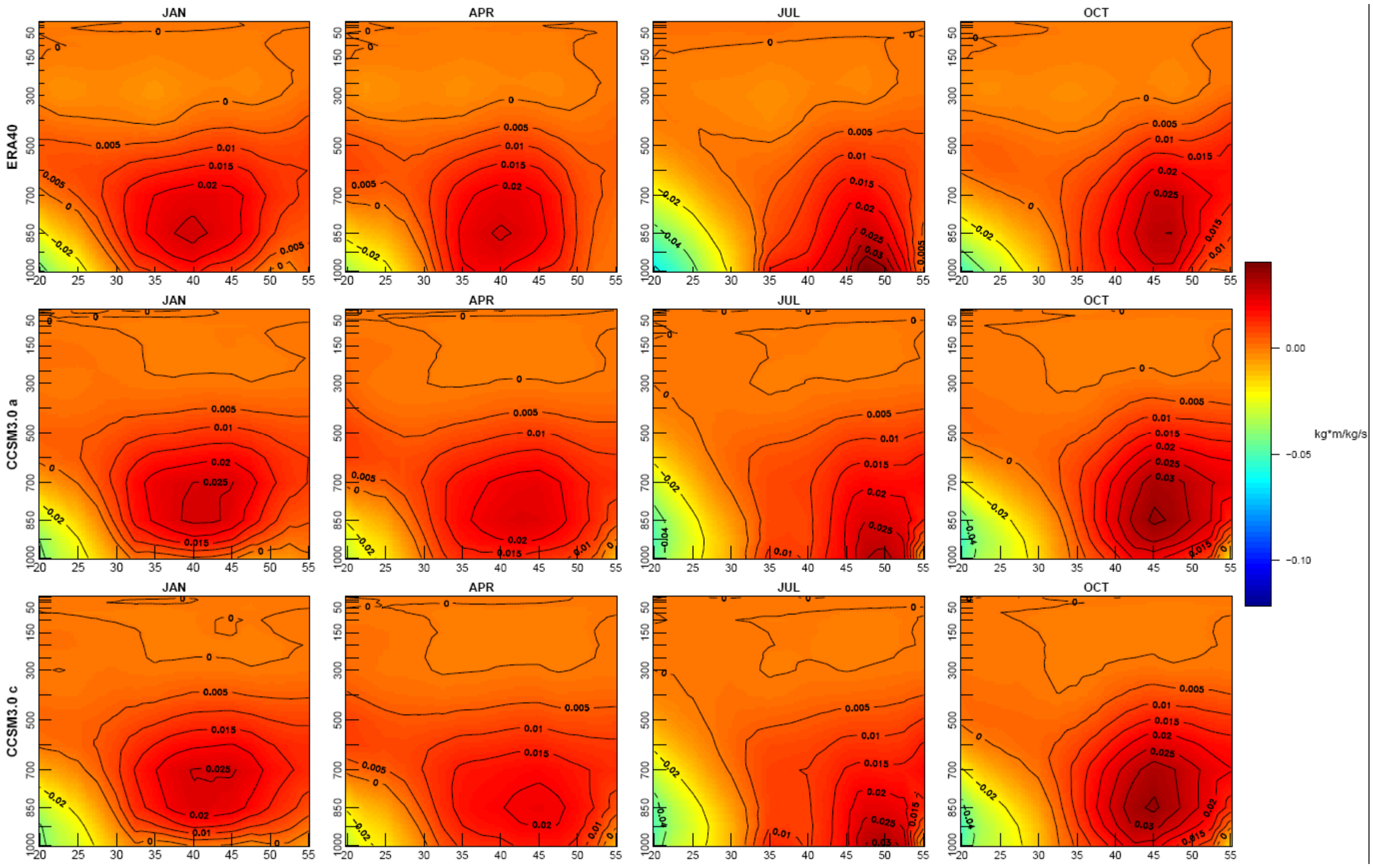

Figure 6.11: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (CCSM3.0) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

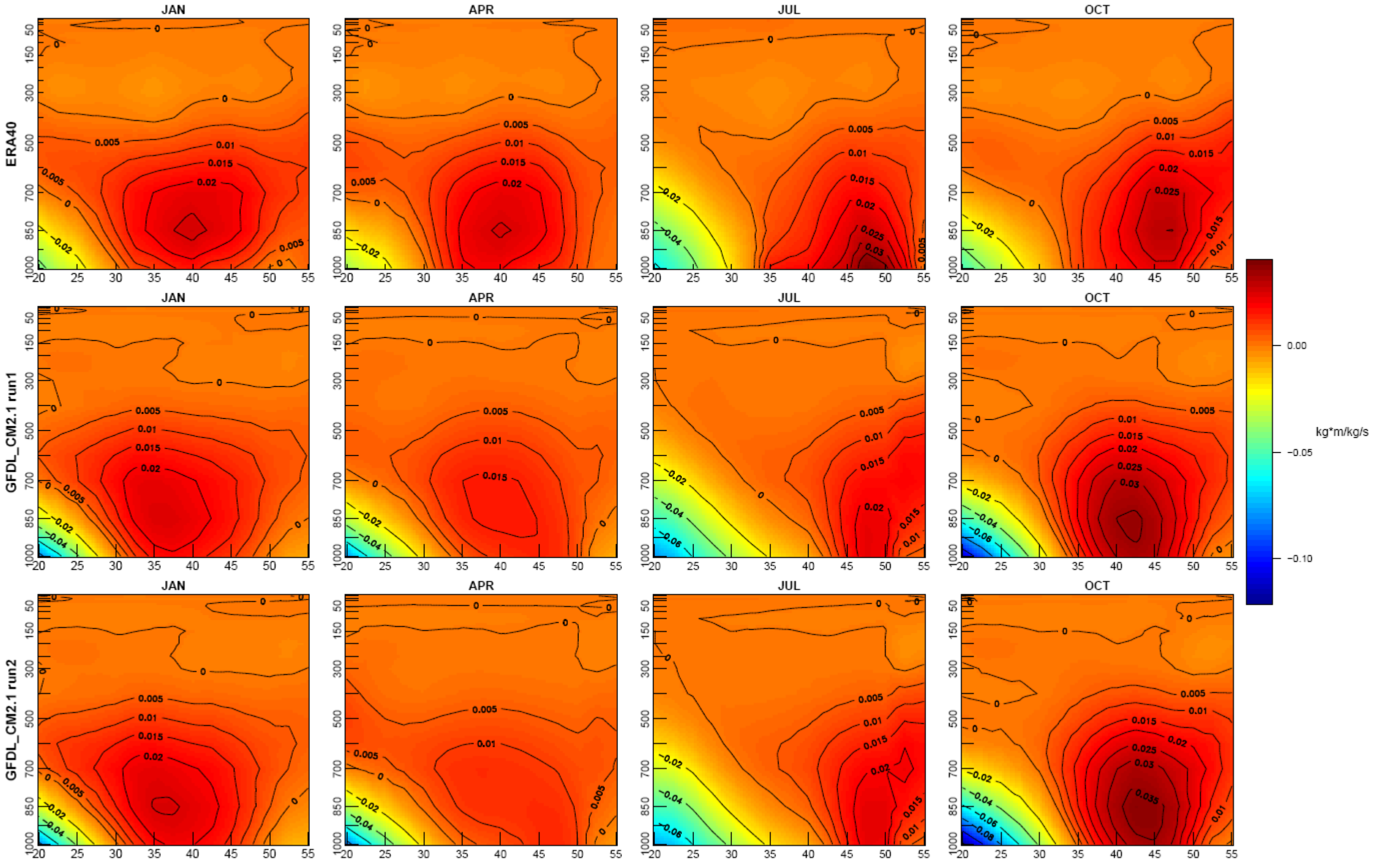

Figure 6.12: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (GFDL2.1) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

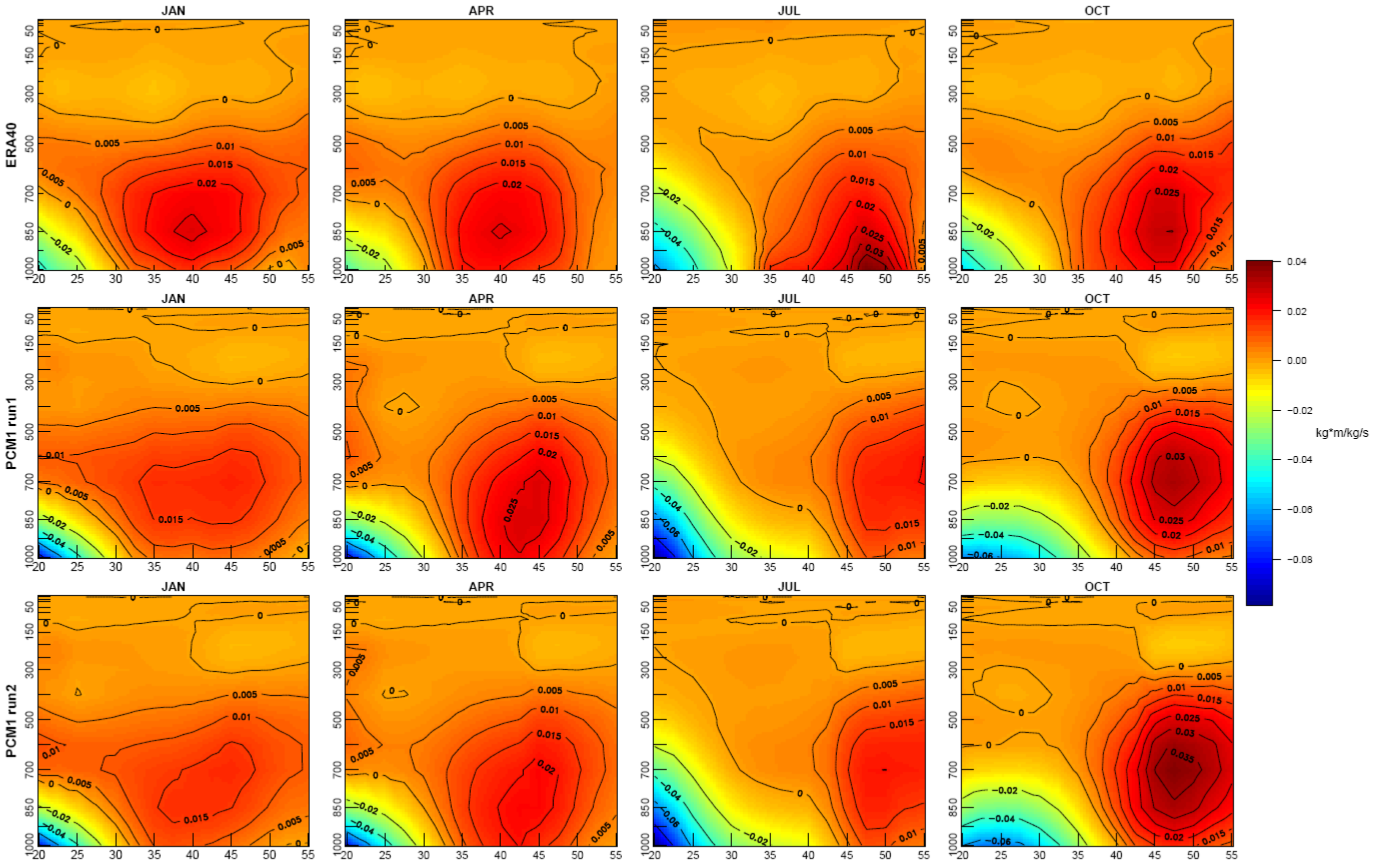

Figure 6.13: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (PCM) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

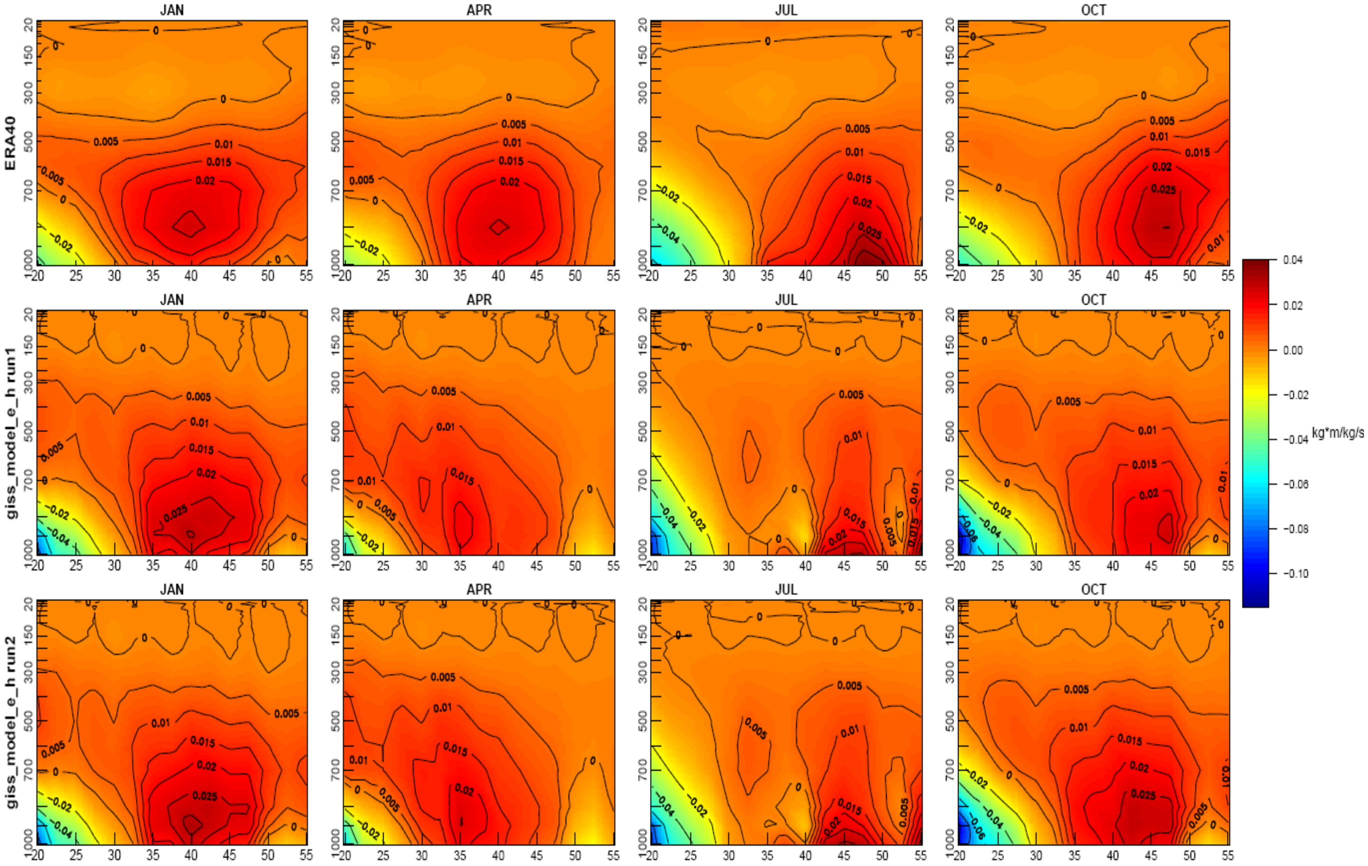

Figure 6.14: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (GISS-EH) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

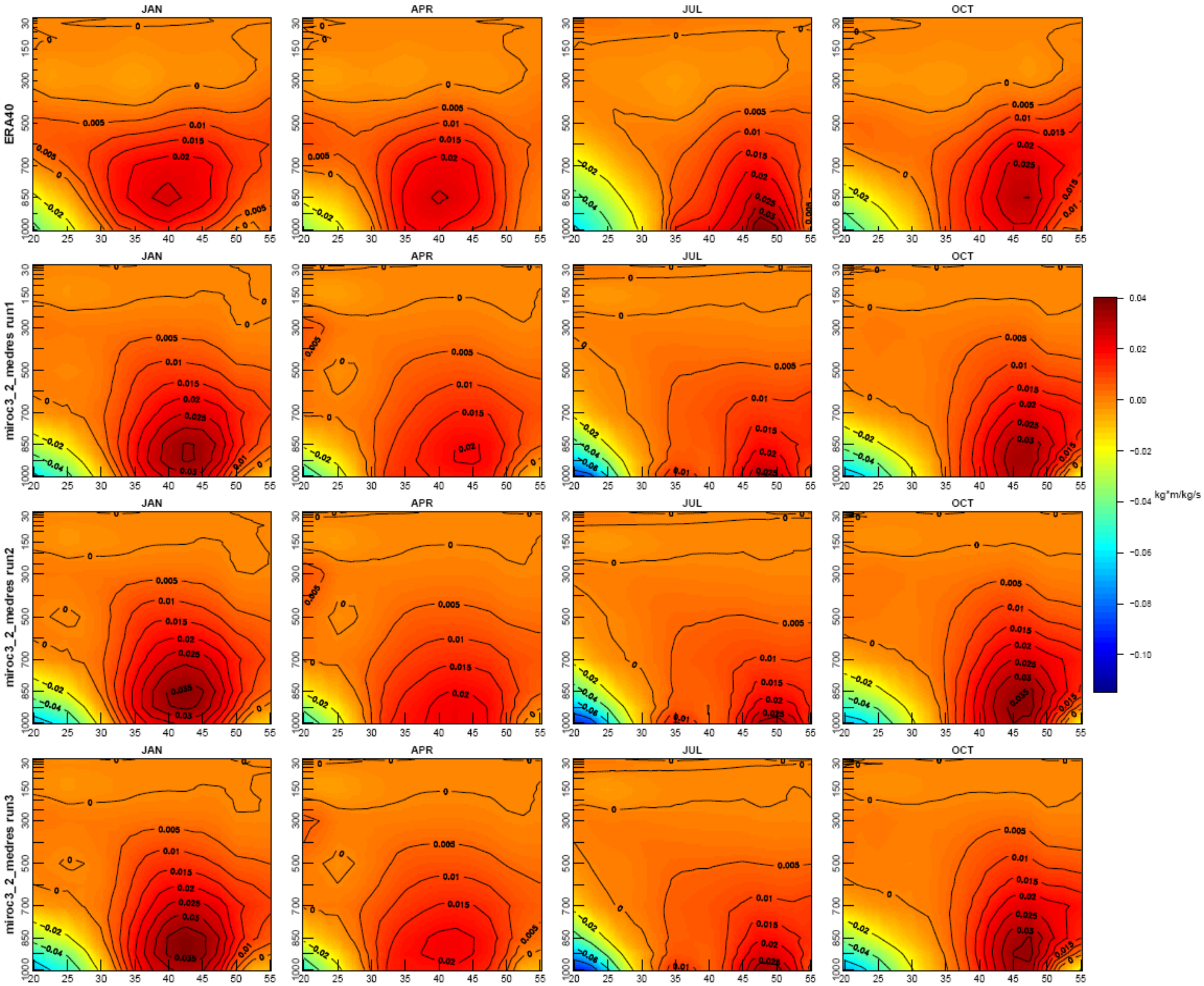

Figure 6.15: Vertical profiles of pseudo moisture fluxes across the western boundary at 130W, comparing model (MIROCmedres) and observed (ERA40 reanalysis) results. (Units are kg m/kg s.). Two runs (using different initializations) are shown for the model.

6.2 Discussion of Results

To facilitate this discussion we summarize the key characteristics of the above results in Tables 6.1 and 6.2 – namely the magnitudes and latitudes of the flux maxima for the pseudo momentum and pseudo moisture fluxes. We consider momentum flux results first.

Table 6.1: Observed and modeled momentum flux characteristics – maximum flux and latitude of the maximum. The MODEL MEAN results are to give a simple overview – it hides large inter-model differences. The first number is the latitude of the maximum (degrees N) and the second number is the maximum flux(m^2/s^2).

MODEL	Jan.	Apr. STJ	Apr. PJ	July STJ	July PJ	Oct. STJ	Oct. PJ
ERA40	**30/650**	**20/710**			**46/230**	**25/240**	**51/650**
CCSM3(a)	25/1220	21/1400		36/230	49/220	25/500	49/920
CCSM3(c)	24/1100	20/1400		37/260		25/500	48/880
GFDL2.1(1)	30/1300	21/700		33/220	50/220		43/930
GFDL2.1(2)	29/1210	22/800		34/220	49/230		45/1000
PCM(1)	27/1500	22/1020	42/810	36/210	55/210		49/820
PCM(2)	26/1620	23/1210		36/220	50/210		51/850
GISS-EH(1)	26/1050	25/1400		32/150	48/220	28/900	46/520
GISS-EH(2)	26/1050	27/1420		33/210	47/220	29/880	42/620
MIROC(1)	41/1010	23/1500		32/230		26/460	47/800
MIROC(2)	41/1210	23/1700		31/220		26/450	46/830
MIROC(3)	42/1100	22/1410		29/260		25/430	47/720
MODEL MEAN	29.6/1226	22.6/1242			49.6/219	26.4/612	46.6/811

For the central winter month (January), when only the sub-tropical jet (STJ) is evident in the observations, all models greatly overestimate the momentum flux (and the speed of the jet), most seriously in PCM. The jet is noticeably too far south in CCSM, PCM and GISS-EH, but reasonably placed in GFDL2.1. In MIROC, because of the simulated jet is so far north, it is likely that the jet that the model simulates is the Polar Jet (PJ). In other words, the MIROC model is simulating incorrect seasonal cycles for both the STJ and the PJ.

By April, GFDL2.1 improves to a realistic representation of the STJ both in terms of position and strength. All other models still seriously overestimate the jet strength, although (with the exception of GISS-EH) they do simulate the position reasonably well. In one simulation, PCM produces a strong PJ, an error in the timing of the PJ seasonal cycle. There is no PJ this early in the observations.

In July, there is no clear STJ in the observations, but all models show a (relatively weak) STJ. MIROC fails to simulate a PJ, another manifestation of this model's failure to get a realistic seasonal cycle for the PJ. The other four models produce a PJ with reasonable location and speed.

In October, both an STJ and a PJ are evident in the observations. Both GFDL2.1 and PCM show reasonable representations of the PJ, although the positions tend to be too far south and the jet is too strong. Neither model has any evidence of an STJ. GISS and MIROC show both jets, but in both models the PJ is too far south (especially in GISS-EH). Jet intensity is good in GISS-EH, but too strong in MIROC. The STJ is well positioned in both of these models, but too strong in MIROC by about 100% and by about 300% in GISS-EH.

To summarize these results, all models generally overestimate the strength of both the STJ and the PJ, often with momentum fluxes being double what is observed and sometimes with model fluxes being as much as four times the observed values. Of the models, GFDL2.1 is superior to the other models. CCSM and MIROC are on a par, although MIROC has serious errors in the timing of the seasonal cycles of both the STJ and the PJ. PCM and GISS-EH are clearly inferior.

Table 6.2: Observed and modeled moisture flux characteristics – maximum fl;ux and latitude of the maximum. The first number is the latitude of the maximum (degrees N) and the second number is the maximum flux(kg m/ kg s).

MODEL	Jan.	Apr. STJ	July PJ	Oct. STJ
ERA40	**40/0.026**	**40/0.026**	**48/0.037**	**47/0.030**
CCSM3(a)	41/0.027	43/0.023	49/0.031	45/0.036
CCSM3(c)	42/0.026	45/0.022	50/0.030	45/0.036
GFDL2.1(1)	36/0.023	40/0.017	48/0.023	42/0.037
GFDL2.1(2)	36/0.026	41/0.012	48/0.023	44/0.038
PCM(1)	42/0.017	43/0.027	55/0.020	47/0.033
PCM(2)	42/0.017	44/0.022	50/0.020	47/0.037
GISS-EH(1)	40/0.030	35/0.022	46/0.027	47/0.026
GISS-EH(2)	40/0.032	35/0.025	45/0.027	44/0.028
MIROC(1)	43/0.036	44/0.021	48/0.026	46/0.026
MIROC(2)	42/0.037	42/0.022	50/0.031	46/0.036
MIROC(3)	42/0.038	41/0.022	49/0.032	46/0.036

Table 6.2 shows results for moisture fluxes. Model fluxes are too weak in all models in April and July. For PCM the moisture flux in October is grossly in error. The flux is too low by almost 50% and the maximum flux is too far north. In January and October, moisture fluxes tend to be too strong. The exceptions are GFDL and PCM in January and GISS-EH in October. Combined with the April and July results, this means that all models are seriously in error in terms of their seasonal moisture-flux cycles.

In addition, in PCM, apart from the flux being much larger than observed, the flux maximum is also too high above ground – the maximum is at around 700 hPa compared with 850 hPa for the observations. This height error in PCM is even more pronounced in July, with the maximum flux at 700 hPa instead of near the surface.

The other major errors are as follows: for CCSM, maximum fluxes are too far north in April; for GFDL, maximum fluxes are not far enough north in January and October; for PCM, maximum fluxes are too far north in April and July; and for GISS-EH, maximum fluxes are not far enough north in April.

In comparing the models, the MIROC model has the fewest deficiencies. CCSM is the next best, then GFDL, GISS-EH and PCM in that order. While there are substantial errors in some seasons, the model moisture fluxes compare much more favorably with the observations than the momentum fluxes. The latter errors are associated, in general, with large errors in the strengths of the sub-tropical and polar jets. As these are in the upper troposphere, where moisture contents are low, they have no significant effect on the moisture fluxes. Still, they may well influence circulation patterns over the western USA, which in turn can effect how much of the moisture that impinges on the land is converted into precipitation. Diagnosing these linkages is a task for the future.

7.0 Significance of Flux Differences

In this chapter we consider the statistical significance of differences between modeled and observed fluxes. We use the Mahalanobis Distance (MD) to quantify model/observed differences, and test to see whether MD values are significantly different from what would be expected by chance due to natural variability (Section 7.1). Most of the calculations use pseudo fluxes, so we conclude (Section 7.2) with a discussion of true versus pseudo fluxes.

7.1 Use of the Mahalanobis Distance

In order to compare two spatial fields (such as a model field and an observed field), and to quantify their degree of similarity, we use some form of distance metric. A simple example of such a similarity index is the Euclidean distance metric, defined by

$$D_{Euclid}(\mathbf{s},\mathbf{o}) = || \mathbf{s} – \mathbf{o} || = \{\Sigma_i (s_i - o_i)^2 \}^{1/2}$$

where **s**=(s_1, …, s_d) and **o**=(o_1, …, o_d) are the model and observed vectors, and ‘d’ is the number of data points (here, 255, corresponding to 15 latitude points by 17 points in the vertical). If we divide by ‘d’ before taking the square root, this is simply the root-mean-square difference between the two fields. (Strictly speaking, D is a **dis**similarity metric, in that larger values of D indicate greater dissimilarity.)

A generalization of the Euclidean distance is the normalized Euclidean distance defined by

$$D_{Norm}(\mathbf{s},\mathbf{o}) = \{\Sigma_i [(s_i - o_i)^2]/Z_i^2\}^{1/2}$$

where Z_i^2 is some form of composite variance. This is a way of giving less weight to grid points that have greater variability. As greater variability is an indication of uncertainty in the mean, inverse variance weighting puts less emphasis on points where the means are more uncertain.

The Euclidean distance is an appropriate metric for isotropic, stationary spatial fields. However, the assumption of isotropy is often not a good one in climate science. Instead, we use the Mahalanobis Distance (MD), a measure that calculates the degree of similarity between two random vectors that have the same mean, @, and covariance structure, @, and thus applies to anisotropic fields. The Mahalanobis distance is defined as

$$D_M(\mathbf{s},\mathbf{o}) = \{ (\mathbf{s} - \mathbf{o}) \Sigma^{-1} (\mathbf{s} - \mathbf{o}) \}^{1/2}.$$

Here Σ is the covariance matrix of the observation-based reanalysis field. If the time variations in **s** and **o** at different latitude-height grid points are uncorrelated, then the Mahalanobis Distance simply reduces to the normalized Euclidean distance. The MD is a more general way to account for uncertainties in the data sets being compared.

We now use the MD to compare model and observed fields. The model simulations are compared to two reanalyses products, the NCEP/NCAR and ERA40 reanalyses. These reanalyses are based on similar observations, but they differ in the models used to process the observational data sets. Errors in the reanalyses arise from observational measurement errors, or biases, as well as errors resulting from the model synthesis. There is an extensive literature on these problems. Nevertheless, for most climate variables, the time-varying spatial patterns are very similar in the two reanalyses. Comparing them gives us a baseline against which to assess differences between model and reanalysis results. If the latter differences are similar to the differences between the two reanalyses, this would mean that the model result was an excellent simulation of the observations.

We concentrate on three particular fields, monthly-mean fields over the 20-year period 1980-1999 for zonal wind, temperature and specific humidity along 230E, from 20N to 55N. Additionally, we calculate the momentum flux, heat flux and moisture flux at 230E. The MD value is computed using ERA40 as the reference reanalysis, and comparing this to the CCSM3.0 20th century model simulations, the CMIP3/AR4 model simulations, and the NCEP/NCAR R1 reanalysis.

The 20-year mean fields, calculated for the meteorological variables of interest, are given as $y_i = (1/n)\sum x_{i,t}$ where n is the number of replicates, I represents location and t represents time. Suppose, $x_{i,t}$ are independent and distributed as $N(\mu_I, \tau_i^2)$ where τ_i^2 is a known measurement error such that $\tau_i^2 = (1/(t-1))\sum(x_{i,t} - y_i)^2$ and $[\mu_1, \mu_2, \ldots \mu_n]^T$ are distributed as multivariate normal with mean constant vector, μ_0, and variance $\Sigma_{\mu 0}$. The variance of the underlying constant mean field, $\Sigma_{\mu 0}$, is estimated by a Bayesian Treed Gaussian Process Model (BTGP; Gramacy and Lee, 2008), and parameterizes the correlation structure using a stationary, anisotropic, nonseparable Gaussian correlation as given by

$$K(x_j, x_k) = \exp\left\{-\sum((x_j - y_k)^2 / d\right\} \qquad (7.1)$$

The nonseparable correlation parametrization fits the range parameter, d, specifying the distance at which two locations exhibit no correlation. A stationary and isotropic covariance structure assumes that the statistical properties of the field do not change under translation and rotation.

To reduce the MD's sensitivity to outliers, we temper the estimated spatio-temporal covariance structure required to calculate the MD value using a basic hierarchical model. For assessing significance we need a threshold value for MD. When the covariance matrix is known, for a Gaussian process the square of the MD is distributed as chi-square with p degrees of freedom. At the 5% significance level the chi-square statistic with $p=255$ degrees of freedom is 293, so only MD values more than $(293)^{1/2} = 17.1$ would be significant at the 5% level (i.e., only values larger than this would indicate a significant difference in the fields). When the correlation matrix is unknown, however, and is estimated using sample data, the resulting distribution of the MD contains more uncertainty and is known to be distributed as Hotelling's two-sample t^2-distribution, a generalization of Student's ***t*** statistic that is used in multivariate hypothesis testing. In our case, the covariance matrix is not estimated by the sampled data, but by BTGP (see equ. 7.1). When the covariance structure is estimated parametrically, the MD distribution is no longer in closed form. We therefore estimate the null distribution using Monte Carlo simulations based on the Gaussian covariance structure parameters. The critical value for significance at the 5% level is still around 17.

Some data points were excluded from the computation of the MD – specifically, those points where values at the lower pressure levels were sometimes below the surface, viz. at 52.5N and 1000hPa, and at

55N and both 1000hPa and 925hPa.

Figure 7.1 shows the 20-year monthly-mean zonal wind field for the ERA40 and NCEP R1 reanalyses, and the CCSM3.0 run a, c and e model simulations. Zonal wind fields are averaged for each January over 20 years, 1980-1999. The Figure also shows, on the right column, the normalized Euclidean distance fields, as defined above, signed as CCSM minus ERA40; and, top of right column, the spatio-temporal inter-annual standard deviation field (SD) for ERA40. The spatio-temporal field accounts for both spatial and temporal variations in the field. The locations with small variation provide heavier weighting between the observation and model differences. Likewise, the locations with large variation weight the differences between model and observation less heavily. As noted previously, the jet in the model is noticeably stronger than in the observations (see the normalized Euclidean distance fields). The jet is also the location of greatest variability (the high variability at the top of the domain is an edge effect, and does not adversely affect the MD values because it is strongly downweighted.)

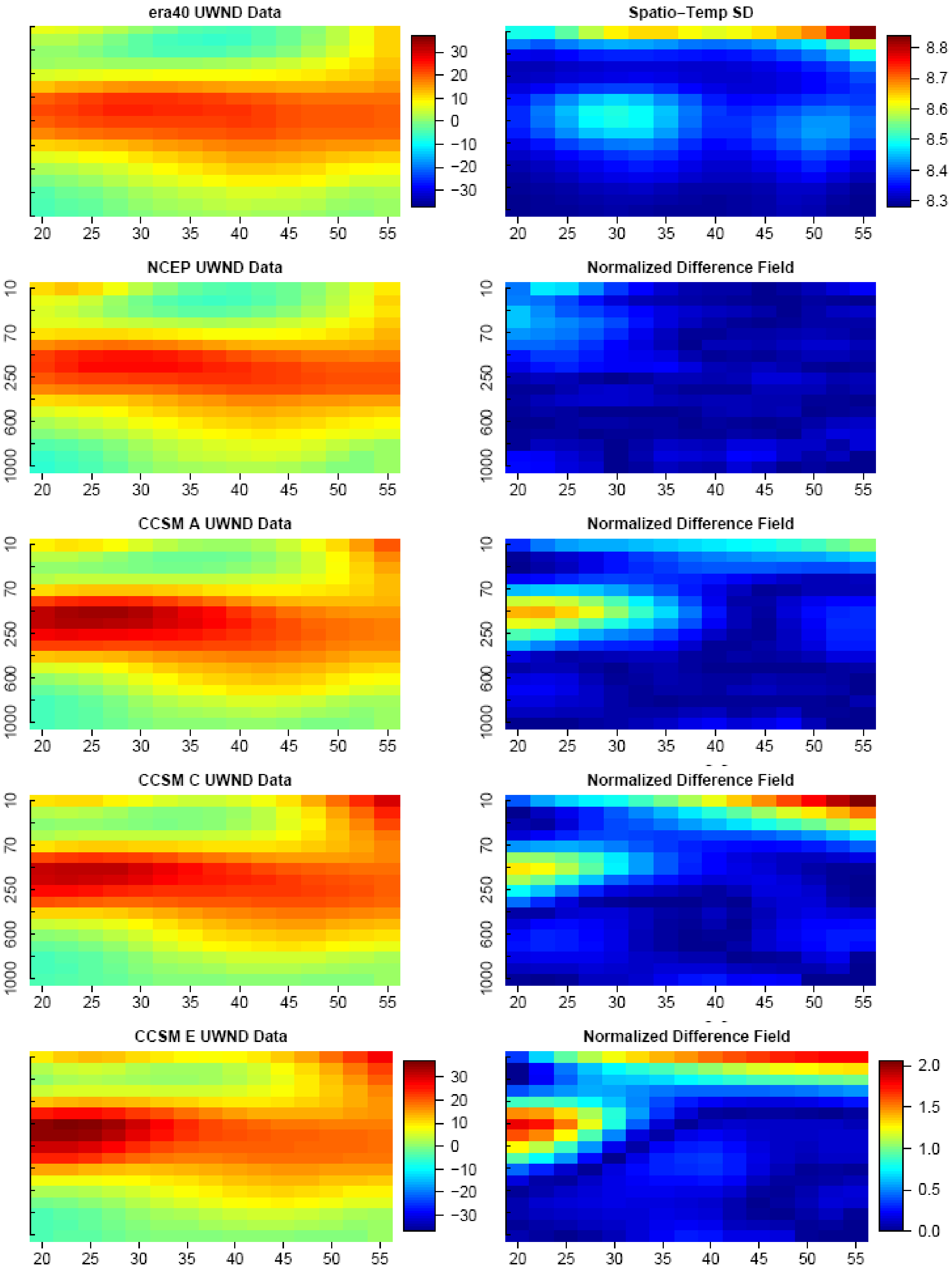

Figure 7.1: 20-year monthly mean zonal wind field for the ERA40 and NCEP R1 reanalyses, and the CCSM3.0 run a, c and e model simulations. Zonal wind fields are averaged for each January over 20 years, 1980-1999.

Figure 7.2 shows MD results for ERA40 versus NCEP and ERA40 versus CCSM runs a, c and e. The MDs conform to the visual similarity between the ERA40 reanalysis and model-simulated mean fields. These results show no significant difference between the reanalysis zonal fields and marginally significant differences between the model and observed fields. These differences arise through differences in the jet speed, which are apparent in Figure 7.1. The largest jet wind speed differences are for run ‘e’, which also gives the largest MD value. The results shown here are for January only, but qualitatively similar results are obtained for the other months.

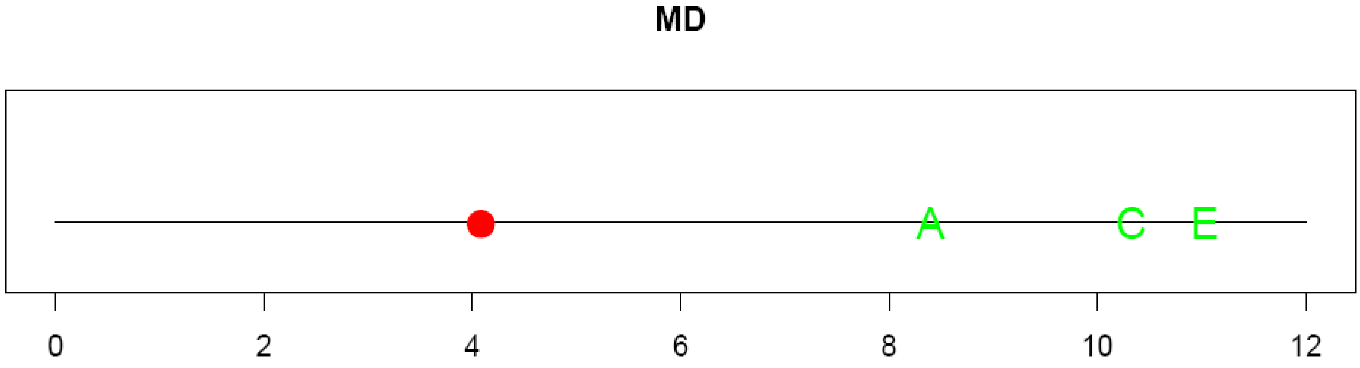

Figure 7.2: Mahalanobis Distance (MD) between the ERA40 and NCEP reanalyses, and CCSM3.0 20th century runs A, C and E corresponding to the fields in Figure 7.1. The MDs are computed with respect to ERA40. NCEP is indicated as a red circle; CCSM simulations are displayed in green.

Figure 7.3 shows the 20-year monthly-mean specific humidity fields for the ERA40 reference reanalyses against the NCEP reanalysis and the CCSM3.0 simulated fields (i.e., the same as Figure 7.1, but for humidity instead of zonal wind). Figure 7.4 shows the corresponding MD results. Figures 7.5 and 7.6 display the analogous plots for the 20-year monthly mean temperature fields. For temperature it is clear that there are significant differences between the model and ERA40 – but this result must be tempered by the fact that ERA40 and NCEP also show marginally significant differences. For water vapor, there are clear differences between the two reanalyses (Figure 7.5, right column, second panel from top). These differences are greater than the differences between the CCSM3.0 model simulations and ERA40, as is clear from both Figure 7.3 and Figure 7.4.

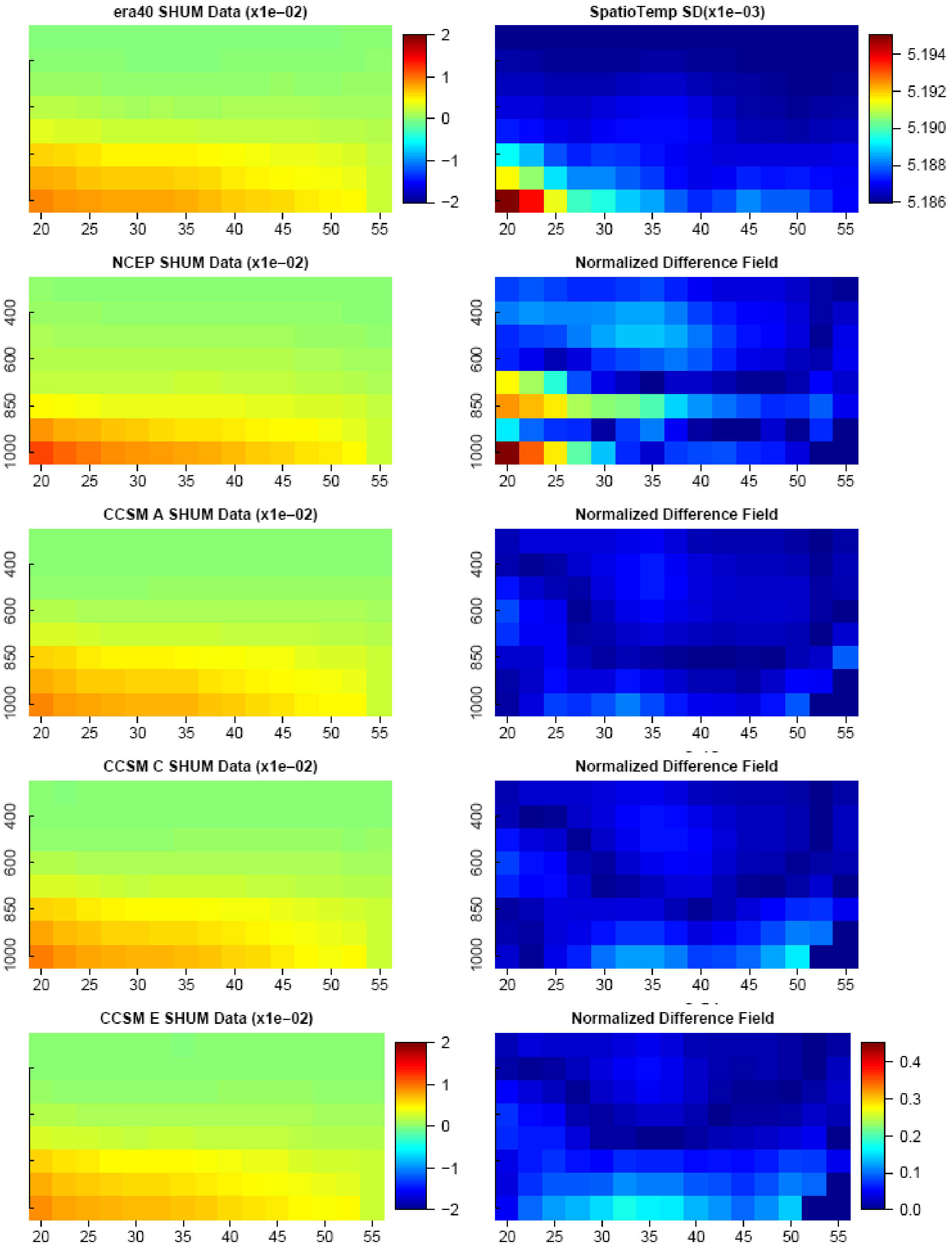

Figure 7.3: 20-year monthly mean specific humidity fields for the ERA40 and NCEP R1 reanalyses, and the CCSM3.0 run a, c and e model simulations. Fields are averaged for each January over 20 years, 1980-1999.

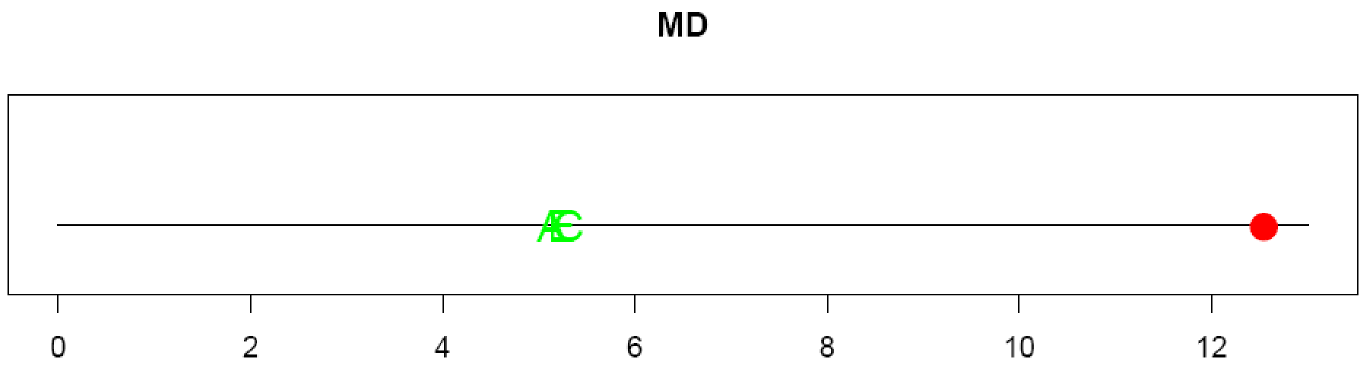

Figure 7.4: Mahalanobis Distance (MD) between the ERA40 and NCEP reanalyses, and CCSM3.0 20th century runs A, C and E corresponding to the fields in Figure 7.4. The MDs are computed with respect to ERA40. NCEP is indicated as a red circle; CCSM simulations are displayed in green.

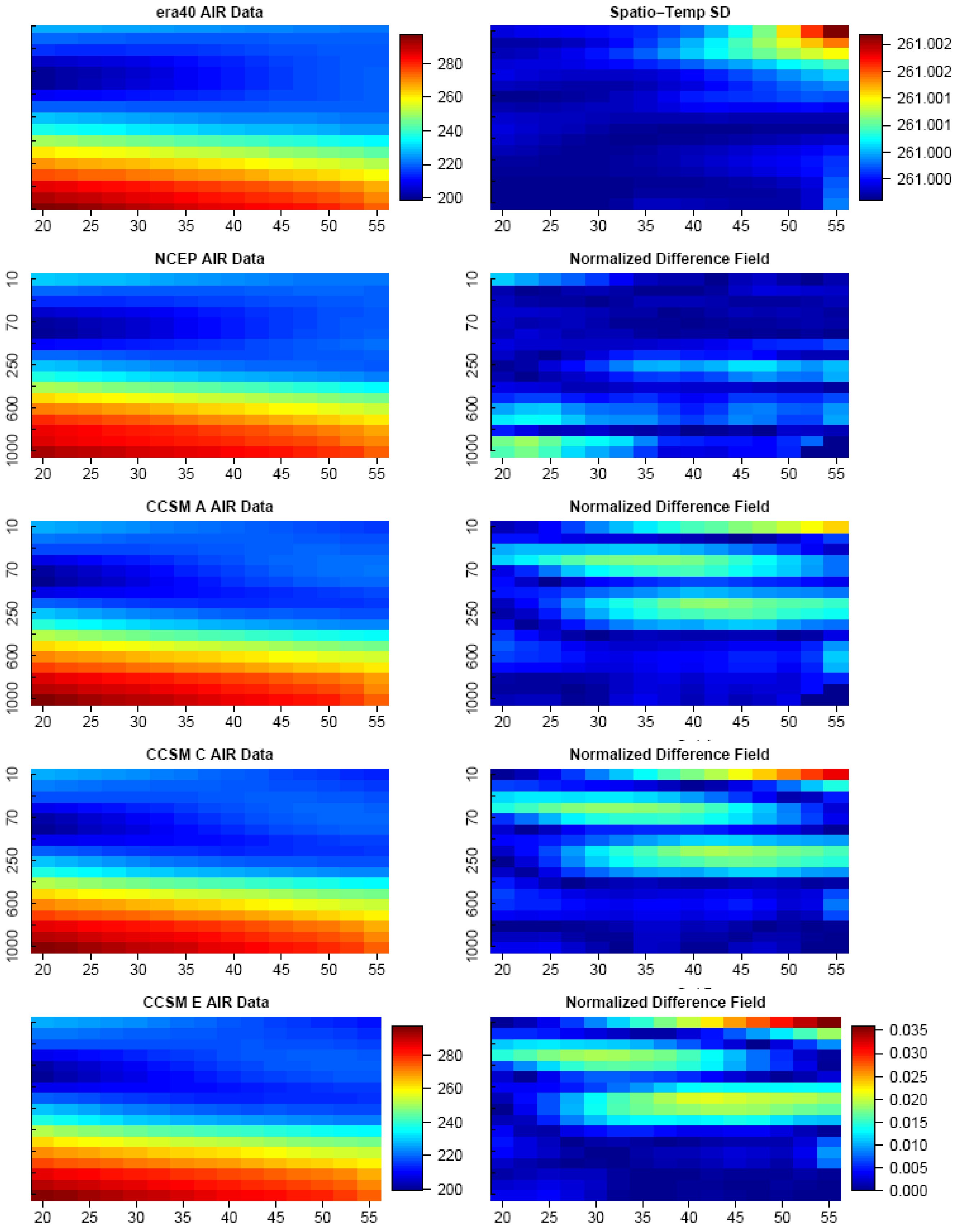

Figure 7.5: 20-year monthly mean temperature fields for the ERA40 and NCEP R1 reanalyses, and the CCSM3.0 run a, c and e model simulations. Fields are averaged for each January over 20 years, 1980-1999.

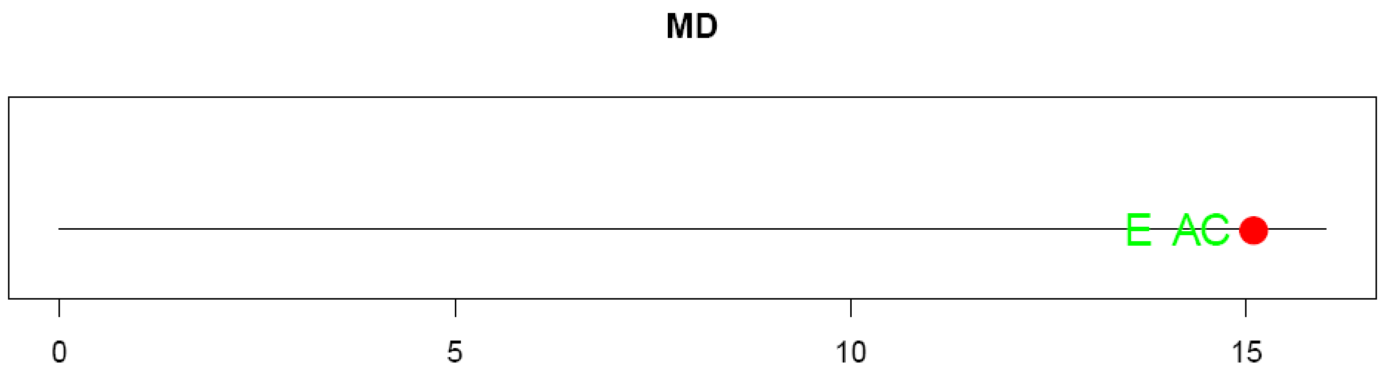

Figure 7.6: MDs corresponding to the fields in Figure 7.5.

The above results show that the MD is a useful new metric for use as a measure of the similarity (or dissimilarity) between spatial fields. In addition to the above examples, we have rigorously tested the MD by simulating spatial fields using a multivariate normal distribution with a prescribed correlation structure to gain understanding of the strengths and weaknesses of the MD as a measure of spatial field similarity.

Ultimately, we want to compare the model simulations of the CMIP3/AR4 global circulation models to the ERA40 and NCEP R1 reanalysis data sets. Most of the GCM data sets are recorded on monthly timescales in the CMIP3 archive, whereas 6-hourly data are required to compute fluxes. We only have 6-hourly data for one model and one run, CCSM3.0, run 'e'.

The flux of variable X at grid point 'i' is defined by $<u_i X_i>$, where u_i is the zonal wind speed and <..> denotes a time average. For cases where 6-hourly data are not available, we have introduced the concept of pseudo fluxes defined by $<u_i><X_i>$. Pseudo fluxes can be calculated with monthly data. We have shown that pseudo fluxes may be used instead of true fluxes for assessing the validity of model fluxes. We now compare the ERA40, NCEP and CCSM3.0 run e data and MDs for momentum, heat and moisture pseudo fluxes.

Figure 7.7 shows momentum fluxes constructed using 6-hourly data, for the NCEP, ERA40 and CCSM3.0 run e simulations in October (left column) together with the normalized difference fields and the spatio-temporal SD field for ERA40.

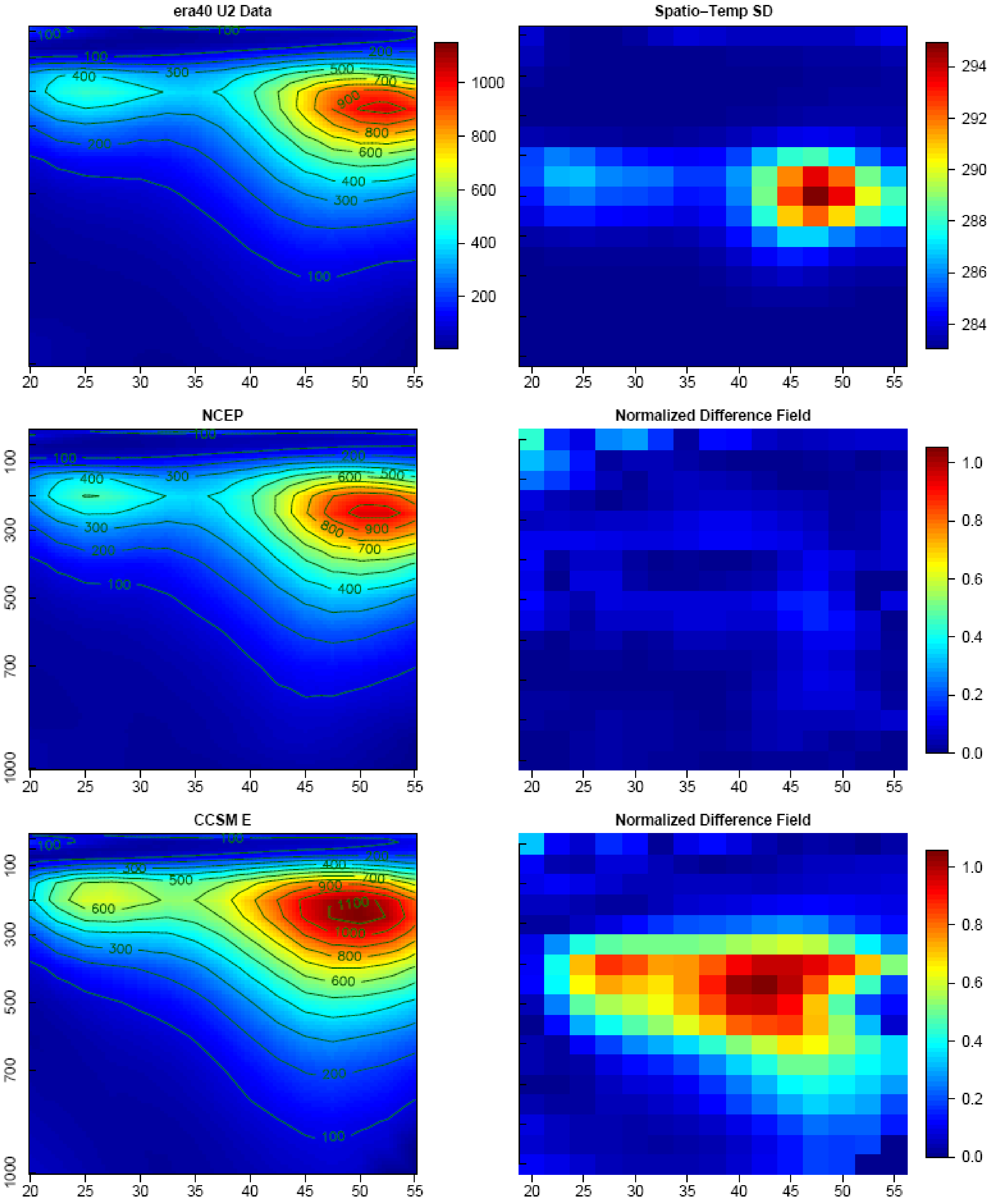

Figure 7.7: True momentum fluxes.

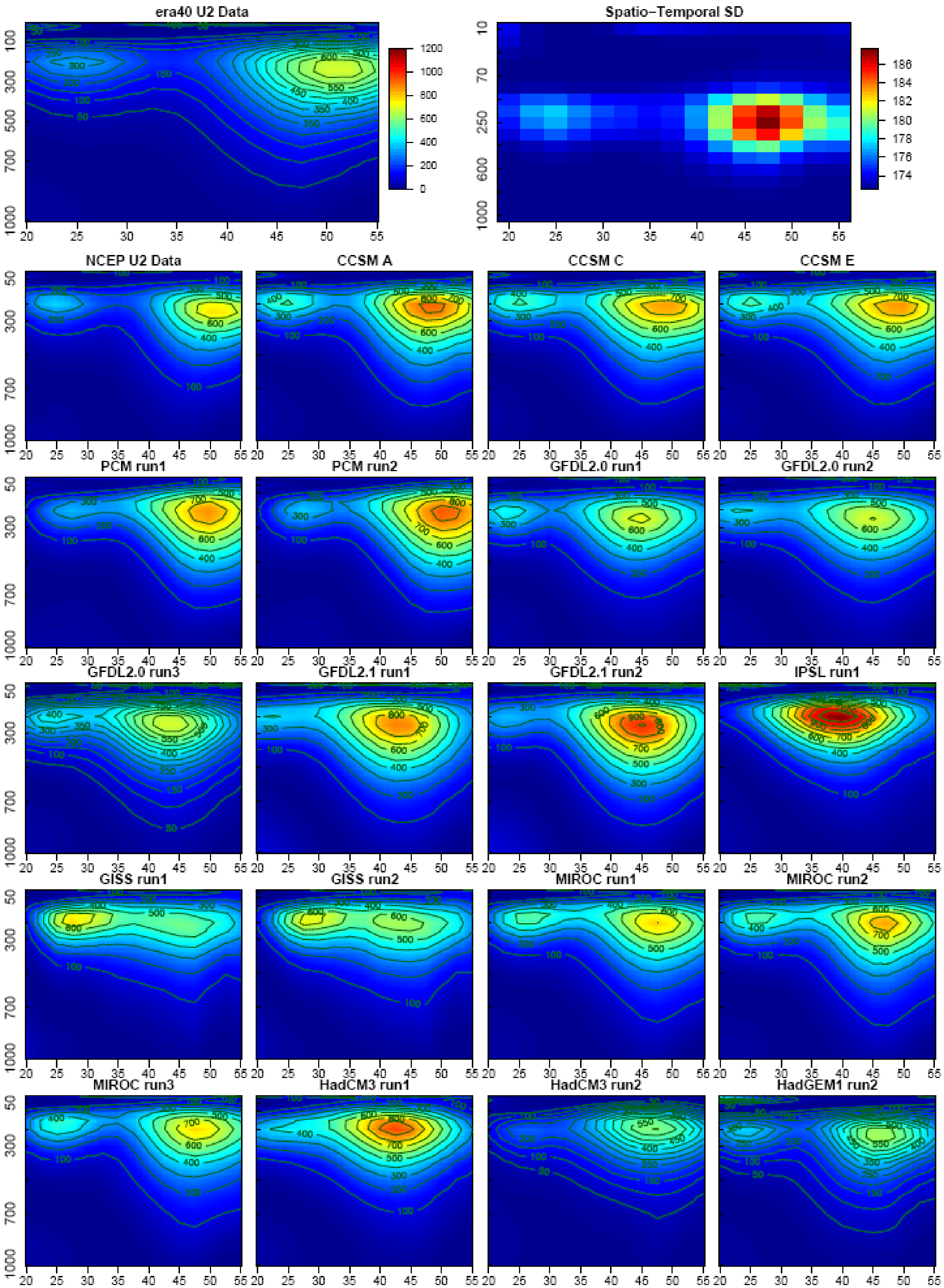

Figure 7.8: See text for details.

Figure 7.8 shows pseudo flux data analogous to Figure 7.7 for the ERA40 reanalysis and the CCSM3.0 run e simulation. The pseudo fluxes are computed using the monthly data; whereas the fluxes in Figure 7.7 are computed using the 6-hourly data. Figures 7.7 and 7.8 provide a comparison between the true flux data and the pseudo flux fields (which are the fields used to compare the CMIP3/AR4 models to the reanalyses data). Figure 7.8 also shows a subset of the CMIP/AR4 20th century simulations, including CCSM3.0 runs a and c, GFDL runs 1 and 2, PCM runs 1 and 2, GFDL2.0 runs 1, 2 and 3, GFDL2.1 run 2, IPSL run 1, GISS runs 1 and 2, HadleyCM3 runs 1 and 2, HadleyGEM1 run2.

Figure 7.9 shows the MD results from Figure 7.7, viz. for ERA40 versus NCEP, and ERA40 versus CCSM3.0 run e. The results are similar to those in Figure 7.2, except that the model shows an even larger difference from the observations (as expected because the flux involves the square of the mean wind speed). Note that the MD magnitudes for the fluxes computed using 6-hourly data closely match the pseudo flux MD magnitudes. The MD values are dependent on the difference between the mean fields, and the spatio-temporal variation in the reanalysis field. Even though these quantities vary on different time-scales, the true flux and pseudo flux MDs approximate each other well for each of the 20-year monthly mean momentum flux fields.

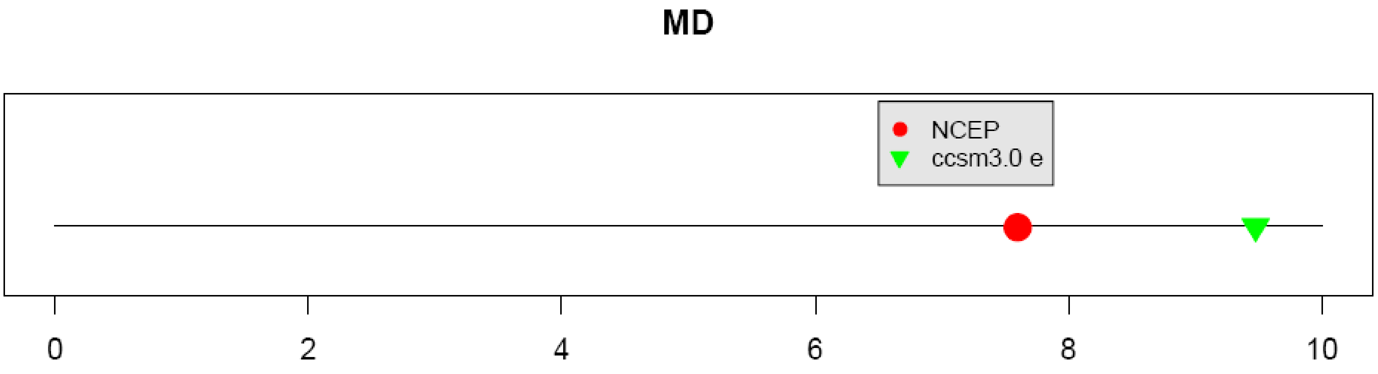

Figure 7.9: MD results for data in Figure 7.7.

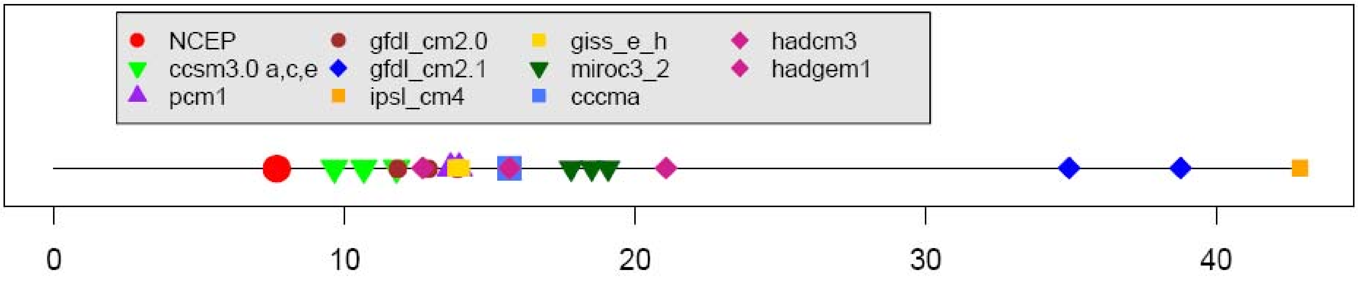

Figure 7.10: MD results for momentum pseudo flux data in Figure 7.8.

Figure 7.10 shows MD results for the momentum pseudo fluxes in Figure 7.8. The model MDs range approximately from 10 to 75. Some of the CMIP3/AR4 model resemble the ERA40 momentum flux field more closely than they resemble the NCEP reanalysis. The GFDL CM2.0, CCSM3.0 and GISS models are more similar to the ERA40 reanalysis. The IPSL model has the largest MD of all the CMIP3/AR4 models included in this plot. Clearly, this model has a significant error, associated with its over-estimate of wind speed in the sub-tropical jet. Similar results are obtained for January, April and October pseudo flux fields.

The corresponding 20-year averaged heat flux results for the month of July are shown in Figures 7.11a (the analog of Figure 7.7) and 7.12a (the analog of Figure 7.8). Figures 7.11b and 7.12b show the computed MDs corresponding to Figures 7.11a and 7.12a, respectively.The July heat flux MDs range from approximately 15 to 40. Again GFDL CM2.0, CCSM3.0 and the Hadley models most closely resemble the ERA40 mean field according to the MDs and the IPSL model lags far behind. The magnitudes of the MDs do vary by month. For example, the April and January MDs for the pseudo fluxes range from only 5 to 25; whereas, the October pseudo flux MD values range from about 8 to 40. In all months, except for July, the NCEP reanalysis has the smallest MD relative to the CMIP3/AR4 models, marking the best resemblance to ERA40.

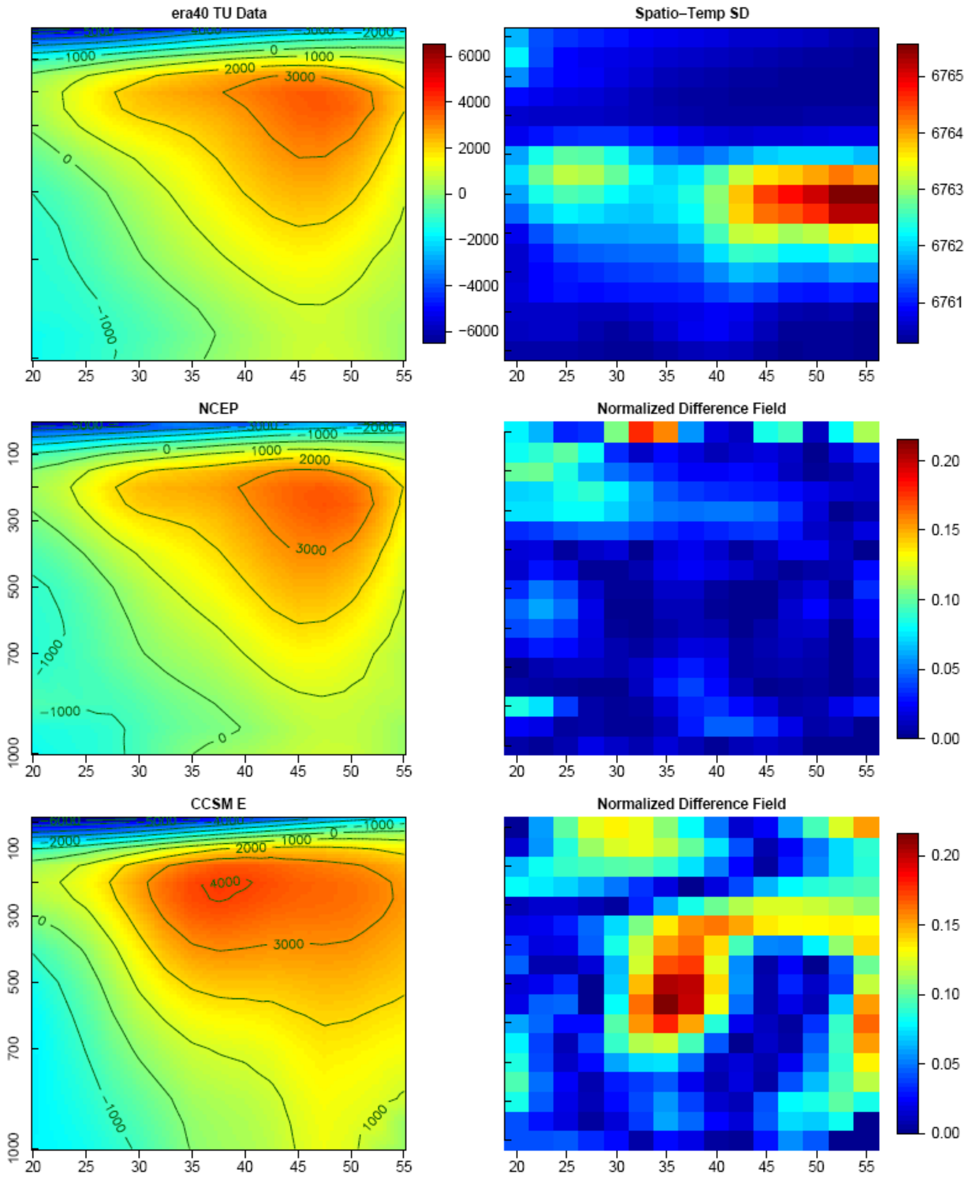

Figure 7.11a: July heat fluxes (cf. Figure 7.7).

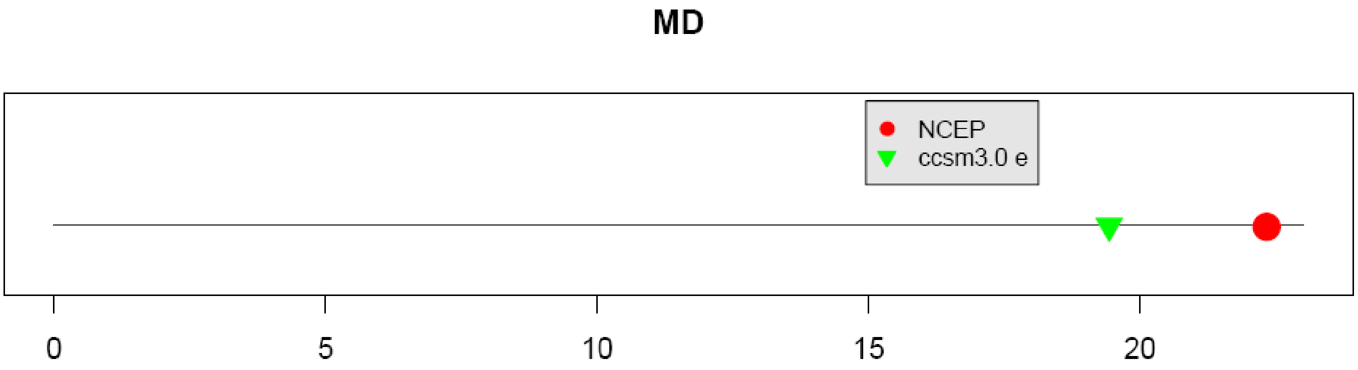

Figure 7.11b: MD values for Figure 7.11a.

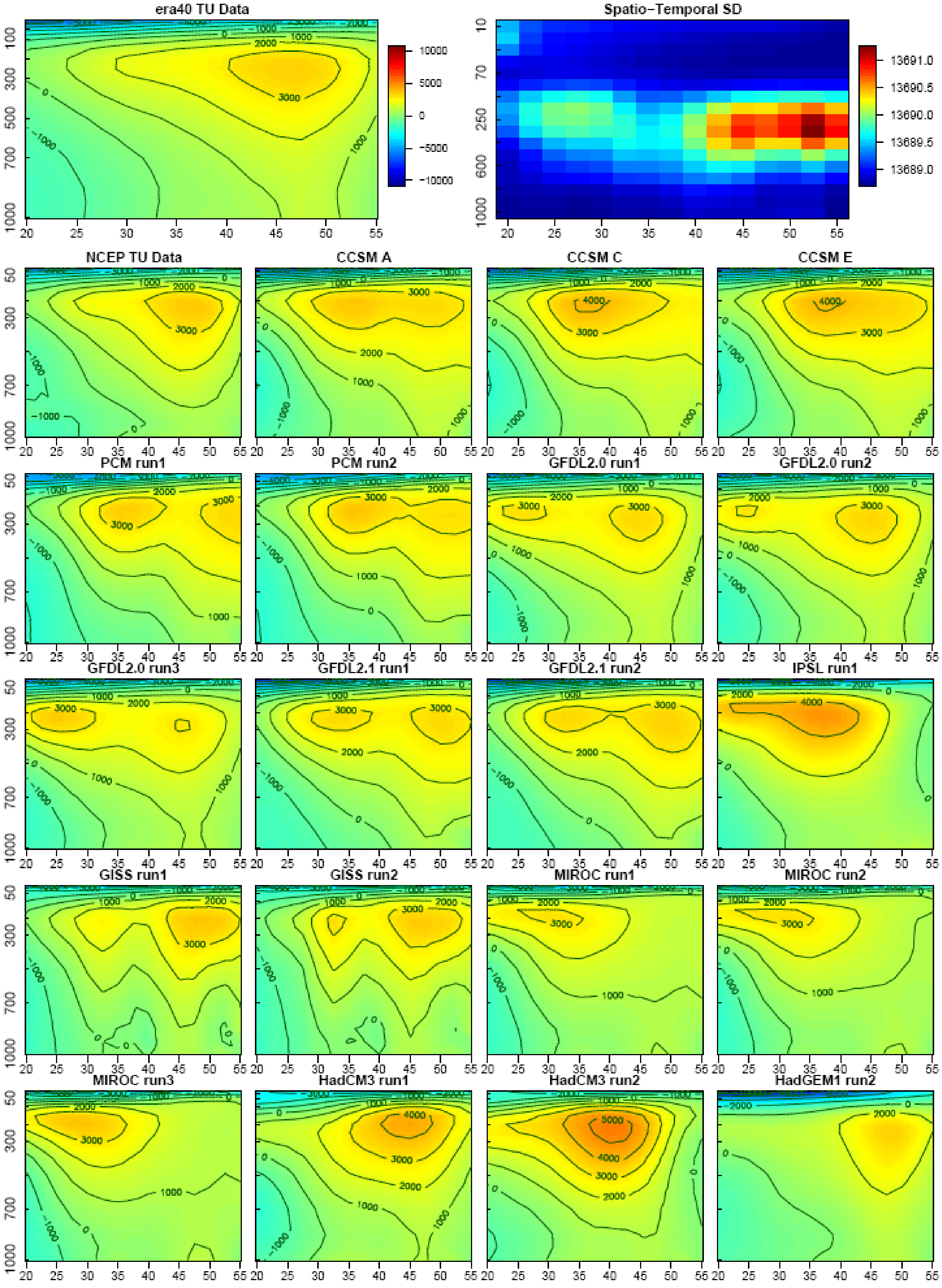

Figure 7.12a: July heat fluxes (cf. Figure 7.8).

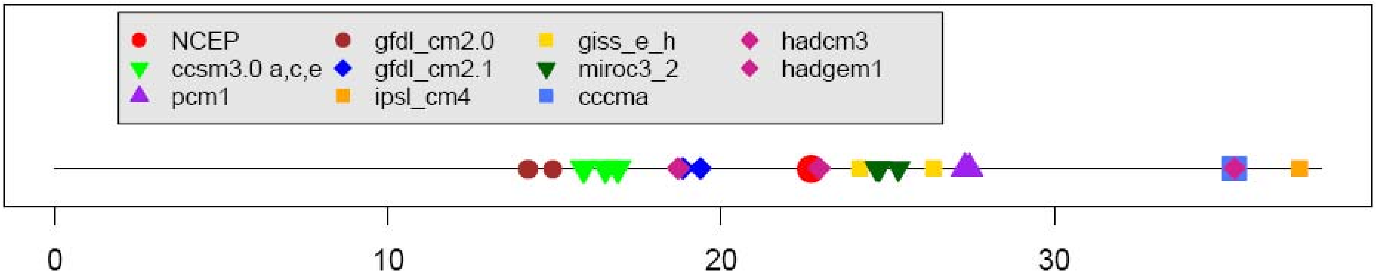

Figure 7.12b: MDs for data in Figure 7.12a.

Figures 7.13a (cf Figure 7.11a) and 7.14a (cf Figure 7.12a) show the July moisture pseudo fluxes for the reanalysis and GCM simulations. Figures 7.13b and 7.14b show the corresponding MDs for moisture fluxes. Many of the simulations match the pattern of the ERA40 moisture flux field, as well as the NCEP reanalyses. The GFDL CM2.0, CCCMA, Hadley and CCSM3.0 simulations are amongst those that produce the smallest MDs, whereas, the GISS MDs are much larger. Visually, the GISS model shows poor pattern correlation.

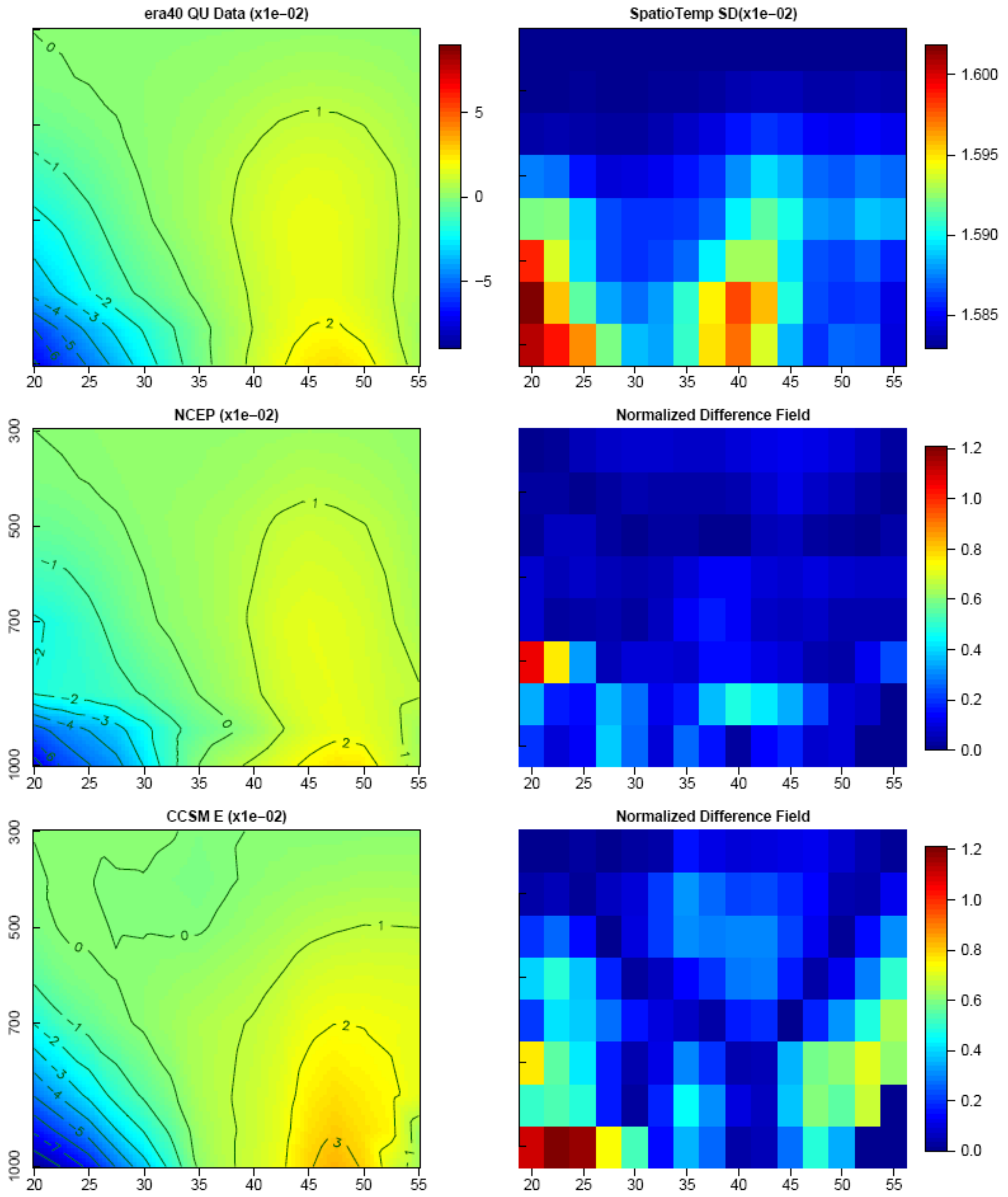

Figure 7.13a: July moisture pseudo fluxes.

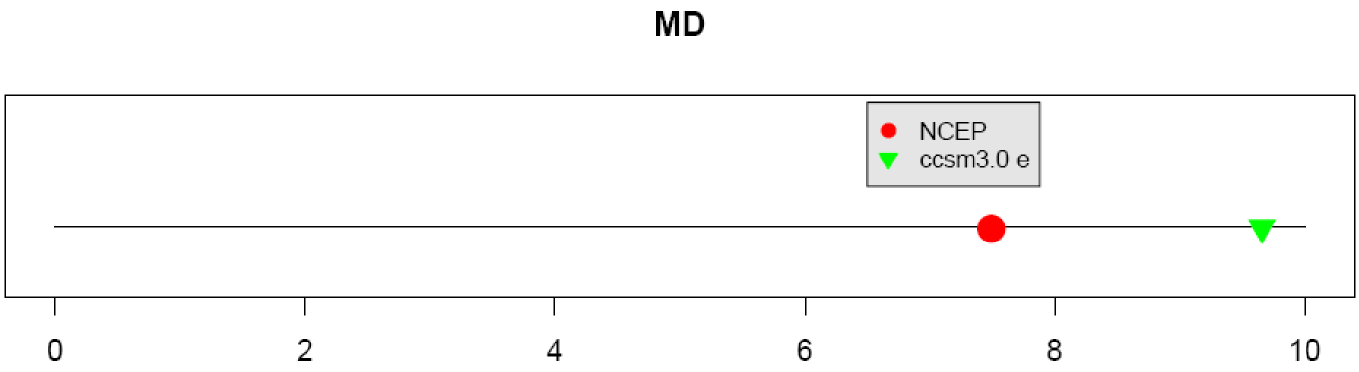

Figure 7.13b: MDs for the data in Figure 7.13a.

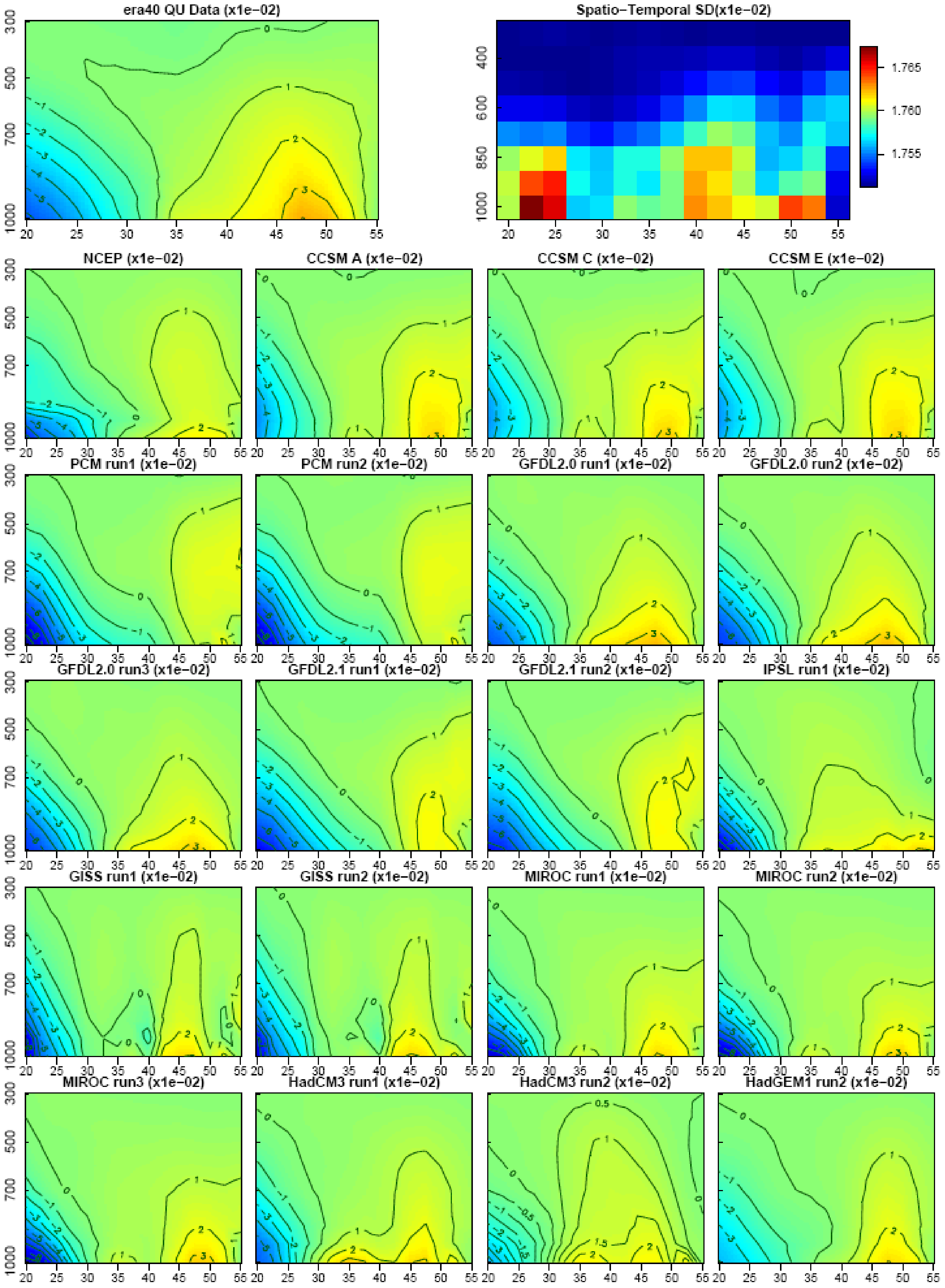

Figure 7.14a: July moisture fluxes.

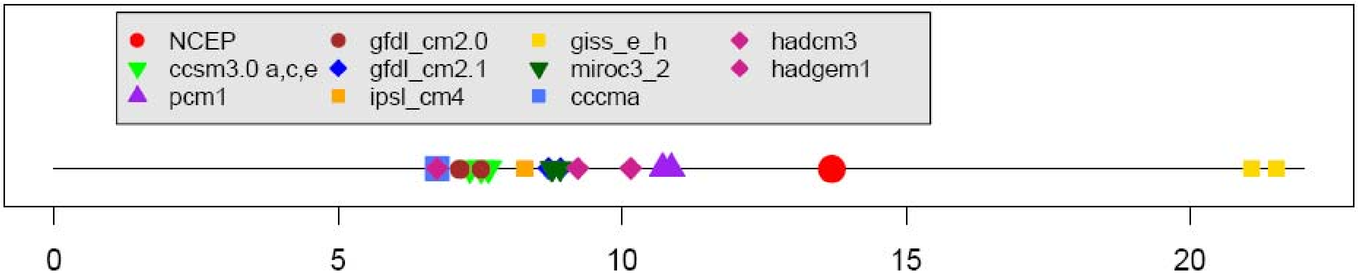

Figure 7.14b: MDs for the data in Figure 7.14a.

7.1 True flux versus pseudo flux results

The NCEP reanalysis moisture flux field, based on 6 hourly data, is a better match to the ERA40 data than the CCSM3.0 run e simulation (see Figure 7.13b). However, the reverse is true for the pseudo-flux fields (see Figure 7.14b). In this case all three CCSM3.0 runs (a, c and e) match the ERA40 reanalysis more closely than the ERA40/NCEP match. Whether this recersal is significant, however, is doubtful. The relatively low MD magnitudes produced by the moisture flux fields mean that none of the moisture flux fields are significantly different, within the variability in the fields themselves. This is as much a result of the variability as it is an indication of model skill. With highly variable fields it is more difficult to detect differences in the mean fiekds. For heat fluxes, however, the magnitudes of the heat flux fields MDs are much larger, by a factor of ten, conveying greater differences with respect to the ERA40 reanalysis variability. We see similar relationships between the moisture flux and pseudo-flux derived MDs for January, April, July and October monthly mean fields.

8.0 References

AchutaRao, K. and Sperber, K.R., 2006: ENSO simulation in coupled ocean-atmosphere models: are the current models better? *Climate Dynamics* **27**, 1–15.

Basist, A.N. and Chelliah, M., 1997: Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis, NCEP operational analysis, and the Microwave Sounding Unit. *Bulletin of the American Meteorological Society* **78**, 1431–1447.

Bengtsson, L. and Shukla, J., 1988: Integration of space and in situ observations to study global climate change. *Bulletin of the American Meteorological Society* **69**, 1130–1143.

Caldwell, P., Chin, H.-N.S., Bader, D.C. and Bala, G., 2009: Evaluation of a WRF dynamical downscaling simulation over California. *Climatic Change* **95**, 499–521.

Christensen, J.H. and Hewitson, B. (Coordinating Lead Authors), together with 15 Lead Authors and 40 Contributing Authors, 2007: Regional Climate Projections. (In) Climate *Change 2007: The Physical Science Basis* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.J. Miller, eds.), Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 847–940.

Giorgi, F. and Mearns, L.O., 2002: Calculation of average, uncertainty range, and reliability of regional climate change from AOGCM simulations via the Reliability Ensemble Averaging (REA) method. *J. Clim.* **15**, 1141–1158.

Gleckler, P.J., Taylor, K.E. and Doutriaux, C., 2008: Performance metrics for climate models. *J. Geophys. Res.***113**, D06104, doi:10.1029/2007JD008972.

Gramacy, R.B. and Lee, H.K.H., 2008: Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association* **103**, 1119—1130.

Gregory, J.M. and Mitchell, J.F.B., 1997: The climate response to CO2 of the Hadley Centre coupled AOGCM with and without flux adjustment. *Geophys. Res. Letts.* **24**, 15, 1943–1946. [doi:10.1029/97GL01930].

Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**, 437–471.

Mearns, L.O., 2004: Proposal for a multiple AOGCM and RCM climate scenario project over North America (unpublished manuscript).

Pawson, S. and Fiorino, M., 1998: A comparison of reanalyses in the tropical stratosphere. Part 1: Thermal structure and the annual cycle. *Climate Dynamics* **14**, 631–644.

Randall, D.A. and Wood, R.A. (Coordinating Lead Authors), together with 11 Lead Authors and 73 Contributing Authors, 2007: Climate Models and Their Evaluation. (In) Climate *Change 2007: The Physical Science Basis* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.

Tignor and H.J. Miller, eds.), Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 589–662.

Rencher, A.C. 2002: Methods of Multivariate Analysis (Wiley series in Probability and Statistics), 2nd edition. Wiley-Interscience. ISBN 0-471-41889-7, p. 76.

Santer, B.D., Hnilo, J.J., Wigley, T.M.L, Boyle, J.S., Doutriaux, C., Fiorino, M., Parker, D.E. and Taylor, K.E., 1999: Uncertainties in observationally based estimates of temperature change in the free atmosphere. *Journal of Geophysical Research* **104**, 6305–6333.

Santer, B.D., Wigley, T.M.L., Simmons, A.J., Kållberg, O., Kelly, G.A., Uppala, S., Ammann, C., Boyle, J.S., Bruggemann, W., Doutriaux, C., Fiorino, M., Mears, C., Meehl, G.A., Sausen, R., Taylor, K.E., Washington, W.M., Wehner, M.F. and Wentz, F.J., 2004: Identification of human-induced climate change in a second-generation reanalysis. *Journal of Geophysical Research* **109**, D21104, doi:10.1029/2004JD005075.

Simmons, A.J. and Gibson, J.K., 2000: The ERA-40 Project Plan. ERA-40 Project Report series No. 1, ECMWF, Reading, U.K., 62 pp.

Takle, E.S., Gutowski, W.J. and 18 others, 1999: Project to intercompare regional climate simulations (PIRCS): description and initial results. *Journal of Geophysical Research* **104**, 19443–19461.

Trenberth, K.E. and Olson, J.G., 1988: An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bulletin of the American Meteorological Society* **69**, 1047–1057.

Trenberth, K.E., Stepaniak, D.P., Hurrell, J.W. and Fiorino, M., 2001: Quality of reanalyses in the tropics. *Journal of Climate* **14**, 1499–1510.

Wigley, T.M.L. 2004: *Input Needs for Downscaling of Climate Data*. California Energy Commission Special Report, Sacramento, CA. 31 pp.

Appendix 1: NARCCAP

NARCCAP (North American Regional Climate Change Assessment Program) is an international program to serve the climate scenario needs of both the United States and Canada. The goal is to investigate the uncertainties in regional scale projections of future climate and to produce high-resolution climate change scenarios. NARCCAP will use different future emissions scenarios, and a number of RCMs (at around 50km resolution) driven by a range of AOGCMs. The study domain covers the conterminous US and most of Canada. NARCCAP also includes a validation aspect through nesting the participating RCMs within reanalyses. RCMs to be used include HadRM3, RegCM, the Canadian regional climate model (CRCM), the NCEP regional spectral model (RSM), and MM5. Candidate AOGCMs include the Hadley Centre HadCM3, NCAR's CCSM3.0, the Canadian CGCM3 (CCCMA), and a GFDL model. High-resolution global time slice experiments based on the GFDL atmospheric model and the NCAR atmospheric model (CAM3) will also be produced and will be compared with the regional model results.